NAMUR
1961

3

# Cybernetica

## CYBERNETICA

est la revue de l'Association Internationale de Cybernétique.
Elle paraît 4 fois par an.

*is the review of the International Association for Cybernetics.*
*It is issued four times a year.*

---

*Prix et conditions de vente — Price and conditions of sale.*

Abonnement annuel — *Yearly subscription :*

| | |
|---|---|
| membres de l'Association | 150,– F. B. |
| *members of the Association* | 150,– F. B. |
| non-membres : | 300,– F. B. |
| *non-members :* | 300,– F. B. |

Par numéro — *Each number :*

| | |
|---|---|
| membres de l'Association | 50,– F. B. |
| *members of the Association* | 50,– F. B. |
| non-membres : | 100,– F. B. |
| *non-members :* | 100,– F. B. |

Toute correspondance concernant la revue est à adresser à l'Association Internationale de Cybernétique, 13, rue Basse Marcelle, Namur (Belgique).

*All correspondence concerning the review is to be sent to the International Association for Cybernetics, 13, rue Basse Marcelle, Namur (Belgium).*

Secrétaire de Rédaction : M. Roger DETRY

# CYBERNETICA

VOLUME IV
N° 3 - 1961

Revue de l'Association Internationale de Cybernétique
Review of the International Association for Cybernetics

NAMUR

# SOMMAIRE
# *CONTENTS*

# Variabilité et spécificité

par François BONSACK,
*Docteur en Philosophie,*
*Ancien Assistant à l'École Polytechnique Fédérale (Zurich)* (1)

C'est pour moi un très grand honneur, sans doute immérité, que d'avoir été appelé à prononcer ici une conférence en séance plénière. Cet honneur, c'est avant tout à M. Couffignal que je le dois ; j'aimerais l'en remercier ainsi que les autres membres du Conseil d'Administration et j'espère ne pas leur donner l'occasion de regretter leur choix.

Je vais parler de choses très élémentaires, que certains d'entre vous connaissent sans doute fort bien, et je m'en excuse auprès d'eux.

Si j'ai décidé d'en parler tout de même, c'est qu'il m'est apparu, au cours de nombreuses conversations, qu'elles ne sont pas connues de tous, même dans le monde de la cybernétique. Il ne me paraît donc pas inutile d'y revenir et d'essayer de voir clair dans certaines notions fondamentales de la théorie de l'information, notions qui ne sont pas toujours exposées avec toute la clarté désirable.

Je vais donc partir de la notion de quantité d'information selon Shannon. On peut l'introduire sommairement ainsi.

On a un *alphabet* comprenant un certain nombre de *signes,* à l'aide desquels on compose un *message.* On se demande comment estimer l'information contenue dans un tel message.

Il est naturel d'exiger que cette quantité d'information croisse linéairement avec le nombre de signes du message, avec sa longueur, que par exemple un message deux fois plus long contienne deux fois plus d'information.

En outre on voit bien, en essayant de traduire un message rédigé à l'aide d'un alphabet de 16 signes, dans un alphabet de 4 ou de 2

(1) Texte de la conférence générale prononcée au 3e *Congrès International de Cybernétique,* Namur, 11-15 septembre 1961.

signes, que le même message requiert beaucoup plus de signes lorsqu'il est traduit dans un alphabet plus pauvre. On en déduira donc naturellement que la quantité d'information transmise par chaque signe est une fonction croissante du nombre de signes que comprend l'alphabet.

Quelle fonction ? Un alphabet de 16 lettres permet d'écrire des messages deux fois plus courts qu'un alphabet de quatre lettres, quatre fois plus courts qu'un alphabet de deux lettres. On est donc conduit au logarithme, puisque $\log_2 16 = 4$, $\log_2 4 = 2$, $\log_2 2 = 1$, et on définira la quantité d'information par symbole

$$I_{\text{par symbole}} = \log_2 n$$

$n$ étant le nombre de signes que compte l'alphabet. Voilà un premier résultat, bien connu de tous.

Il faut cependant bien se rendre compte comment on l'a obtenu : on n'a pas opéré sur des messages ordinaires, rédigés dans un langage naturel, comme le français ou l'anglais. Ce serait trop compliqué. On se fait donc un modèle simplifié, où l'on considère des « messages » qui sont simplement des suites de lettres disposées plus ou moins au hasard et qui, à première vue, ne signifient rien. Mais on définit la quantité d'information comme s'il s'agissait de véritables messages, signifiant quelque chose. Cette procédure est parfaitement légitime, dans certaines limites. Mais il ne faut pas oublier que c'est sur un modèle qu'on opère, et non sur de véritables messages.

Mettons ce modèle en pleine lumière : les « messages » avec lesquels on opère sont des suites de signes et rien d'autre. Ces suites de signes sont formées par un processus aléatoire, par exemple par des jets de dés, ou par des tirages de boules dans une urne, de cartes dans un jeu brassé. A chaque signe est attribuée une certaine probabilité ; si la suite est assez longue, on peut espérer, avec une certaine sécurité, que la fréquence relative des signes s'approchera de leur probabilité.

Nous allons donc considérer maintenant uniquement le modèle : une machine produisant certaines suites de signes selon certaines lois de probabilités, une machine à fabriquer des suites aléatoires de signes.

Mais ici, dans ce modèle, le terme de « quantité d'information » n'a plus aucun sens, puisqu'il n'y a plus à proprement parler d'information. Il y a une certaine grandeur du modèle qui correspond à ce que, dans des messages signifiants, on peut appeler « quantité d'information », et qui, dans le cas où tous les signes sont équiprobables, est donnée par la formule

$$H_{\text{par symbole}} = \log_2 n$$

Cette grandeur que Shannon a appelée « entropie » et qui, on ne le répétera jamais assez, n'a dans le modèle rien d'une quantité d'information, est maintenant une mesure de la *variété* des suites de signes produites par la machine-à-produire-des-suites-aléatoires-de-signes, de leur diversité, ou une mesure de la *variabilité,* de la richesse de la production de la machine, de la variabilité du « type » de « messages » produits par la machine. Plus les « messages » sont divers, variés, plus l'entropie, la variabilité est grande ; plus au contraire les « messages » sont uniformes, monotones, plus leur entropie-variabilité est petite. (Pour éviter toute confusion, je précise que j'utiliserai le terme de « variabilité » lorsqu'il s'agira d'un *type de messages,* le terme de « variété » lorsque je considérerai une *collection* supposée réalisée).

Ce modèle s'avère très fécond. On peut en particulier passer à des signes de probabilités inégales ; on obtient alors la fameuse formule

$$H_{\text{par symbole}} = - \Sigma\, p_i \log p_i$$

On peut aussi tenir compte des liaisons entre les signes. On se rapproche alors considérablement des langues naturelles (où les signes ne sont ni équiprobables, ni indépendants) de telle sorte qu'on pense, à première vue, pouvoir redonner à la variabilité sa signification primitive de quantité d'information.

Voilà où l'on arrive en suivant une première voie — voie que j'appellerai objective parce qu'on y examine la production réalisée, objective, d'une machine à produire des suites aléatoires de signes, production qu'on peut examiner sans se soucier de signification ou d'information proprement dite.

Mais on peut aussi suivre une voie différente, que j'appellerai subjective parce qu'elle interprète l'entropie non pas comme la variété en soi d'un certain ensemble de « messages », mais comme notre *incertitude* quant à un certain événement. On quitte le domaine de l'être pour celui de la connaissance, on passe d'une variété en soi à une variété pour nous, à la variété des messages que nous croyons pouvoir recevoir.

Cette façon d'envisager les choses est peut-être plus accessible ; le modèle y est peut-être plus proche de la réalité et l'entropie peut-être plus immédiatement interprétable comme une quantité d'information.

En effet, on peut envisager des processus aléatoires — tirages

de cartes ou de boules — et se poser des questions quant à l'information qu'on possède ou qu'on fournit sur le résultat de ces ces processus aléatoires. Par exemple, on peut effectuer un tirage de cinq boules dans une urne contenant, en quantités égales, des boules rouges et des boules blanches et se demander combien on transmet d'information sur le résultat de ce tirage en dévoilant le nombre de boules rouges tirées (sans préciser l'ordre dans lequel elles ont été tirées).

Nous reviendrons un peu plus loin sur cette conception subjective de l'information, mais je voudrais auparavant bien préciser le sens où j'entends « subjectif » et « objectit ».

Il s'agit d'une subjectivité dans un sens très spécial, plus exactement d'une relativité à la situation, au point de vue. En ce sens, un aspect perspectif d'un objet est subjectif parce qu'il dépend du point de vue où l'on se place ; suivant le point de vue, on verra un cercle comme un cercle ou comme une ellipse plus ou moins aplatie.

Ici, je parle de conception subjective de l'entropie parce qu'elle dépend non pas seulement des caractéristiques de l'objet, de l'événement, mais encore de ce que sait le sujet. Lorsque j'apprends quelque chose sur un objet, cet objet ne se modifie pas, en soi ; mais la connaissance que j'en prends se modifie. Mon incertitude au sujet d'un événement peut se modifier ; elle peut diminuer si l'on me fournit des informations ; l'événement se modifie alors *pour moi*, mais non pas *en soi* ; c'est une modification que j'appelle donc subjective.

Elle est encore subjective en ce sens qu'elle varie suivant ce que sait tel ou tel sujet. Supposons, par exemple, qu'on examine l'information apportée par une certaine nouvelle. D'après la définition subjective, la quantité d'information n'est pas la même pour un récepteur qui ne l'attend pas, qui est dans une ignorance, dans une incertitude totale quant à cette nouvelle, que pour un récepteur qui soupçonne déjà quelque chose de précis, qui a donc avant la la réception de l'information une incertitude moins grande. Pour un récepteur qui connaît déjà la nouvelle, celle-ci n'apporte plus d'information, car il n'y a déjà plus aucune incertitude avant la réception de l'information.

Il y a donc une certaine subjectivité, en ce sens que la quantité d'information dépend de la situation de connaissance dans laquelle se trouvait le récepteur avant de recevoir la nouvelle ; mais cette subjectivité n'a rien de la subjectivité du poète ou du passionné ; étant donnée telle situation de connaissance, l'information est

univoquement déterminée ; elle sera la même pour tous les individus dans la même situation, quels que soient leurs goûts ou leur tempérament. Elle est moins subjective au sens courant que *relative à la situation.*

Je ne vais pas choisir entre la conception objective et la conception subjective de l'information. Personnellement, je trouve la conception objective plus satisfaisante et plus générale ; il est en particulier très facile de retrouver la conception subjective en partant de la conception objective. D'autre part, la conception objective ouvre des perspectives du côté de ce qu'on a quelquefois appelé la « nég-entropie structurale », l'entropie d'un certain type de structures qui n'ont rien d'informationnel. Mais je dois reconnaître que la conception subjective présente certains avantages d'exposition et peut-être moins de dangers de confusions.

L'erreur fréquente que je voudrais ici dénoncer n'est d'ailleurs pas particulière à la conception subjective ou objective : on la retrouve dans les deux conceptions. Elle consiste à identifier simplement la quantité d'information à l'entropie, à définir la quantité d'information par exemple à l'aide de la formule $-\Sigma p_i \log p_i$ et à s'en tenir là. Ce n'est pas seulement faux parce que l'entropie est une grandeur du modèle alors que la quantité d'information est une grandeur du domaine représenté — car il pourrait y avoir correspondance assez rigoureuse entre le modèle et ce qu'il représente. C'est plus grave : l'entropie n'est pas la grandeur du modèle qui correspond à la quantité d'information, c'est une autre grandeur, très proche, mais non identique, qui lui correspond. Shannon ne commet pas d'erreur dans son exposé, mais il ne souligne nulle part qu'après le premier chapitre, l'entropie cesse d'être une mesure de la quantité d'information, bien qu'il utilise lui-même quelque chose d'autre.

Si l'on n'est pas suffisamment conscient de ce fait, on risque de se heurter à des paradoxes que l'on n'arrive pas à débrouiller. J'en parle en connaissance de cause, puisque j'ai buté pendant plusieurs années sur de tels paradoxes, avant de découvrir où était l'erreur. C'est dans l'espoir d'éviter de pareils écueils à quelques-uns que j'ai choisi de traiter ce sujet devant vous.

Ces paradoxes se manifestent surtout lorsqu'on aborde le problème de la transmission avec bruit. En voici un exemple.

Prenons des messages formés de deux signes, 0 et 1. Ces deux signes ne sont pas équiprobables : 0 a la probabilité 3/4 et 1 la probabilité 1/4.

Calculons l'entropie d'un tel message à l'aide de la formule

$$H_{\text{par symbole}} = -\Sigma\, p_i \log p_i$$

On a

$$H_{\text{par symbole}} = -(3/4 \log_2 3/4 + 1/4 \log_2 1/4) = 0{,}811$$

Introduisons un bruit, par exemple 1/4 d'erreurs (une erreur étant un 0 pour un 1 ou un 1 pour un 0) à la transmission. On a le schéma suivant

$$3/4\ 0 \left\langle \begin{array}{lll} 3/4 \text{ corr.} & 9/16 & 0 \\ 1/4 \text{ err.} & 3/16 & 1 \end{array} \right.$$

$$1/4\ 1 \left\langle \begin{array}{lll} 3/4 \text{ corr.} & 3/16 & 1 \\ 1/4 \text{ err.} & 1/16 & 0 \end{array} \right.$$

On a donc, après perturbation, $9/16 + 1/16 = 5/8$ de 0 et $3/16 + 3/16 = 3/8$ de 1. L'entropie du message perturbé est

$$H_{\text{par symbole}} = -(3/8 \log_2 3/8 + 5/8 \log_2 5/8) = 0{,}955$$

L'entropie est donc passée de 0,811 à 0,955 ; elle a augmenté. C'est tout à fait normal si l'on interprète l'entropie comme une variabilité : le bruit a introduit une nouvelle source de variabilité, on s'attend donc à ce que celle-ci ait augmenté.

Mais la quantité d'information, elle, a bien évidemment diminué : le bruit, les perturbations n'augmentent pas la quantité d'information, bien au contraire.

Que conclure de cette contradiction ?

On ne peut en conclure qu'une chose, c'est que l'entropie n'est pas la grandeur qui doit représenter la quantité d'information. Il faut donc se mettre à la recherche d'une autre grandeur, qui ne nous mène pas à de semblables paradoxes mais qui explique cependant qu'on ait pu prendre tout d'abord l'entropie-variabilité pour une bonne mesure de la quantité d'information.

Que fait l'expéditeur d'un message ?

Il a à sa disposition une collection de messages possibles d'une certaine étendue ou richesse mesurée par une certaine entropie-variabilité. Par exemple, cette collection sera l'ensemble des suites de signes d'une certaine longueur qu'on peut composer à l'aide des touches d'un téléscripteur.

Mais il ne veut pas envoyer n'importe quel message, n'importe quelle suite de lettres. Il choisit *un message particulier* dans cette collection ; c'est ce message particulier qu'il veut envoyer, et au-

cun autre. C'est d'autre part ce message particulier et aucun autre qui transmettra telles informations au destinataire, informations qu'il désire avoir ou que l'expéditeur désire lui communiquer.

Il semble donc que l'information liée à un message soit liée à sa spécificité, au fait qu'il est ce message et aucun autre. Et si un brouillage en fait un message quelconque, le rend méconnaissable, il n'apportera plus aucune information, même si son entropie-variabilité est toujours très élevée.

Essayons de voir cela un peu plus clairement sur un exemple particulier.

Prenons des messages de 2 signes, 0 et 1, équiprobables. Transmettons-les avec 1/4 d'erreurs. Leur entropie-variabilité n'aura pas changé, puisqu'il y aura autant de 0 qui deviendront des 1 que de 1 qui deviendront des 0. On aura donc à la réception de nouveau des messages composés de 0 et de 1, équiprobables, et qui ne se distingueront en rien des messages non perturbés. Mais quelque chose a changé, quelque chose à quoi doit être liée la quantité d'information : ce ne peut être que le fait que la spécificité des messages choisis par l'expéditeur a diminué.

Supposons que l'expéditeur transmette un grand nombre de fois le même message et examinons ce que ce message est devenu après transmission. Si la transmission se fait sans bruit, ce message donnera toujours le même message : la spécificité du message et l'information qu'il transmet éventuellement sont donc conservés. Mais si la transmission est perturbée, ce même message donnera naissance, à la réception, à une collection de messages d'une certaine variété : dans l'un, il y aura une erreur à la première lettre et à la cinquième, dans l'autre, à la troisième seulement, etc.. Plus les perturbations seront nombreuses, plus la variété de cette collection augmentera et moins grande sera la quantité d'information réellement transmise.

Examinons encore un autre cas. Supposons que j'aie convenu avec mon correspondant que seule une lettre sur deux soit significative, que le reste soit un remplissage quelconque. Dans ce cas, à mon message significatif correspondra un grand nombre de messages différant par le remplissage, mais identiques pour les lettres significatives. Ici encore, la quantité d'information par symbole est nettement moins grande que si tous les symboles avaient été significatifs, et en même temps la spécificité du message a diminué, puisque le message unique a été remplacé par toute une collection de messages équivalents.

Ces considérations suggèrent une solution à notre paradoxe.

On a deux collections : la collection de messages dont on dispose avant le choix et la collection des messages qu'on a sélectionnés parmi ceux de la première collection. Plus la première collection est étendue, diverse, plus un message choisi pourra contenir d'information ; cela, nous l'avons déjà établi dans la première partie. La quantité d'information est donc une fonction croissante de l'étendue de la collection avant le choix. Mais c'est une fonction décroissante de l'étendue de la collection choisie : plus cette collection est large, moins le choix est spécifique, moins grande est la quantité d'information. Nous allons mesurer l'étendue de ces deux collections par leur entropie-variété et définir la *spécificité* de la collection choisie — cette spécificité qui est la mesure de la quantité d'information — par la différence entre ces deux entropies

$$\text{Spécificité} = H_0 - H_1$$

$H_0$ étant l'entropie de la collection offerte à l'expéditeur et $H_1$ étant l'entropie-variété de la collection choisie.

La spécificité est donc une mesure de la particularité, de l'originalité d'une certaine sous-classe restreinte par rapport à une classe plus large, à un référentiel. Cette spécificité est d'autant plus grande que la sous-classe est plus restreinte ou que le référentiel est plus étendu. Par exemple, la sous-classe des individus à yeux bleus a une certaine spécificité par rapport au référentiel des hommes en général ; elle a une spécificité moins grande par rapport aux habitants des pays nordiques ; par contre, la sous-classe des individus à yeux bleus et à cheveux bruns a une spécificité plus grande que celle groupant *tous* les individus ayant les yeux bleus, quelle que soit la couleur des cheveux.

Cette notion de spécificité permettra-t-elle de lever notre paradoxe ? Mènera-t-elle à des résultats raisonnables ?

Essayons de l'appliquer à la transmission avec bruit. L'expéditeur avait à sa disposition une certaine collection de messages d'entropie-variété $H_0$. Parmi ces messages, il en choisit un seul. La collection n'a plus aucune variété ; le nombre de possibilités équiprobables étant égal à 1, l'entropie, qui s'obtient en prenant le logarithme de ce nombre de possibilités, est égal à 0 ; $H_1$ est donc ici nulle. Dans ce cas, l'entropie-variété de la collection offerte à l'expéditeur est donc bien une mesure de la quantité d'information, puisqu'on n'a rien à lui soustraire.

Mais, après pertubation, l'entropie-variabilité du message ne sera plus nulle ; à un certain message émis correspondra une certaine collection de messages possibles à la réception. On doit donc

estimer la spécificité de cette collection correspondant à un message particulier par rapport à l'ensemble des messages possibles à la réception (correspondant à des messages quelconques à l'émission). En d'autres termes on cherche à calculer la spécificité de ce qu'est devenu le message par rapport à l'ensemble des messages possibles à la réception.

Dans l'exemple de tout à l'heure (3/4 de 0 pour 1/4 de 1, 1/4 d'erreurs à la transmission), nous avions calculé l'entropie des messages à la réception. Elle était égale à 0,955. Nous devons soustraire de cette entropie, l'entropie de la collection issue d'un message particulier ; cette entropie est égale à

$$-(3/4 \log_2 3/4 + 1/4 \log_2 1/4) = 0{,}811$$

La spécificité de cette collection par rapport à l'ensemble des messages possibles à la réception est donc égale à

$$0{,}955 - 0{,}811 = 0{,}144 \text{ bit par symbole}$$

Un quart d'erreurs fait donc perdre plus des 4/5 de l'information, ce qui peut paraître surprenant à première vue, mais s'explique par le fait qu'on perd non seulement de l'information parce que telle lettre a été substituée à telle autre, mais encore parce qu'on ignore où ces substitutions ont été faites, parce que rien ne distingue un *a* issu d'un *a*, d'un *a* issu d'un *e* ou de n'importe quelle autre lettre. En supprimant simplement des lettres et en signalant l'endroit où une lettre a été supprimée, 1/4 de suppressions ne diminuerait l'information que d'un quart.

Passons à l'autre exemple, celui où les 0 et les 1 sont équiprobables ; l'entropie-variabilité par signe est donc égale à 1.

Introduisons un nombre considérable d'erreurs : supposons que la moitié des signes soient perturbés.

Nous avons déjà vu que l'entropie-variabilité était la même à l'émission et à la réception. Quelle est la spécificité des messages reçus ? La variété de la collection issue d'un message donné vaut

$$-(1/2 \log_2 1/2 + 1/2 \log_2 1/2) = 1$$

L'entropie-variété de cette collection est donc égale à celle de tous les messages à la réception. La spécificité est donc nulle, et la quantité d'information aussi : en toute rigueur, plus aucune information n'est transmise. Il y a bien encore la moitié des signes qui sont corrects, mais comme rien ne permet plus de les reconnaître, ils ne peuvent être utilisés. On obtiendrait une aussi bonne « transmission » en coupant la ligne et en composant le message

en jouant à pile ou face ; dans ce cas aussi, il y a en moyenne un signe sur deux qui coïncide avec le signe correspondant de n'importe quel original.

Cet exemple nous permet de préciser encore un peu le sens du mot spécificité. Tant que la collection des messages issus d'un message particulier est plus restreinte que la collection de tous les messages possibles à la réception, cette collection garde une certaine particularité, une certaine spécificité ; les collections issues de messages différents se distinguent encore les unes des autres, il y a encore des traces, des restes de la structure du message émis. Mais lorsque le message a été complètement désorganisé, la collection issue d'un message particulier n'a plus rien de particulier, plus rien de spécifique, plus rien de typique ; elle ne se distingue plus de la collection issue de n'importe quel autre message, toutes les traces de la structure primitive ont disparu, la spécificité tombe à zéro. On peut caractériser brièvement ce fait en disant que le désordre complet n'a pas d'odeur, qu'il ne trahit plus rien de son origine, parce que tous les désordres absolus se ressemblent.

On arrive d'ailleurs exactement à la même formule, pour la quantité d'information, dans l'interprétation subjective. L'entropie y est interprétée comme une incertitude quant au message. Si la transmission est absolument fidèle, on passe de l'incertitude maximale à une incertitude nulle ; l'information apportée par le message est donc égale à l'incertitude au départ.

Mais si le message est brouillé, sa réception diminue bien un peu l'incertitude, mais ne la supprime pas totalement. La quantité d'information apportée par le message est alors égale à la diminution d'incertitude qu'il a permise, à la différence entre l'entropie avant et l'entropie après la réception du message

$$I = H_0 - H_1$$

Si l'incertitude est restée ce qu'elle était, l'information reçue est nulle. Tout ceci s'accorde fort bien avec la notion intuitive d'information.

* * *

Arrêtons nous un instant pour faire le point.

Nous avons défini :

— d'une part, une *entropie,* qui est une mesure de la *variabilité* d'un certain type de messages, ou de la *variété* de la collection de messages de même type à laquelle il appartient ;

— d'autre part, une *spécificité*, qui est une mesure de la particularité, de l'originalité d'un message ou d'un certain groupe de messages par rapport à l'ensemble des messages de même type. Cette spécificité s'exprime par une différence entre deux entropies-variabilités. C'est cette spécificité, et non l'entropie-variabilité, qui représente dans le modèle la quantité d'information.

Quels sont les rapports, les ressemblances et les différences entre ces deux notions ?

L'un des rapports est évident : la spécificité est une différence entre deux variabilités.

Nous avons déjà fait allusion à un autre rapport : lorsque la sous-classe dont on estime la spécificité ne contient plus qu'un seul élément, la spécificité de cet élément (exactement de ce type d'élément singulier) est égale à l'entropie-variabilité du référentiel (puisqu'on lui soustrait une entropie nulle).

Mais il y a aussi des différences.

La différence essentielle, c'est que la spécificité est *relative*, alors que l'entropie-variabilité est *absolue*. La variabilité comporte une échelle absolue, du fait qu'on n'a besoin d'aucune convention, d'aucun choix de référence pour situer le zéro de cette échelle : il y a entropie ou variabilité nulle lorsqu'il n'y a qu'une possibilité, lorsque le type ne comprend qu'une seule forme ou, dans l'interprétation subjective, s'il y a certitude. La spécificité est au contraire relative ; il faut toujours préciser par rapport à quel référentiel on la mesure. Un message est moins spécifique par rapport à un référentiel lui-même hautement spécifique que par rapport à un référentiel peu spécifique. Un texte français par exemple aura une spécificité moins grande par rapport à l'ensemble des textes français que par rapport à l'ensemble de tous les textes possibles dans toutes les langues, ou par rapport à des suites aléatoires de signes sans aucune liaison.

Cette relativité de la spécificité, le fait qu'elle soit définie comme une différence de deux termes, la rapproche d'une part d'une différence de potentiel, comme le font fort justement remarquer les Yaglom dans leur petit livre sur la Probabilité et l'Information, d'autre part de la *probabilité*, qui est elle aussi relative : elle est constituée par un rapport entre deux ensembles d'événements, l'ensemble des cas favorables et celui de tous les cas possibles. Ce rapport, ce quotient est devenu ici une différence, parce qu'on est passé des nombres aux logarithmes.

Il faut également s'arrêter un instant à la question du signe ; certains auteurs, M. Brillouin en particulier, ont identifié la quantité d'information à une entropie négative, à une nég-entropie.

Cette solution me paraît un peu trop sommaire. Le problème se présente en réalité de la manière suivante.

Tout d'abord, nous le verrons tout à l'heure, la définition — dans la théorie moléculaire — de l'entropie a subi au cours de son histoire une profonde modification : on est passé d'une définition probabiliste selon Boltzmann à une définition statistique selon Planck ; au lieu de définir, comme Boltzmann, l'entropie à partir du logarithme d'une probabilité, on l'a définie à partir du logarithme d'un nombre de complexions. Ce faisant on a, à mon avis, inversé le signe de la relation de l'entropie avec la probabilité, car le nombre de complexions, c'est l'inverse de la probabilité d'une complexion, si toutes les complexions sont équiprobables. L'entropie de Shannon est définie, selon Planck, à partir du nombre de complexions (ou, si l'on veut, à partir du logarithme négatif d'une probabilité). Il y a déjà là une question de signe à laquelle il faut prendre garde.

Mais il y en a une autre.

La question des signes se pose surtout à propos de la *variation* d'entropie ; ce qui importe, c'est de savoir si l'entropie augmente ou diminue.

Or il semble bien que l'entropie augmente lorsque la quantité d'information diminue — c'est ce qui a amené M. Brillouin à son identification de la quantité d'information à une nég-entropie. Comment la chose se présente-t-elle si l'on introduit la notion de spécificité ?

La spécificité comprend deux termes : l'un positif, l'entropie du référentiel ; l'autre négatif, l'entropie du sous-ensemble dont on estime la spécificité.

La réponse est donc très claire : si la variation d'entropie touche le premier terme, il n'y a aucune inversion de signe ; une augmentation de l'entropie du référentiel augmente la spécificité. Mais si la variation touche le second terme, il y a alors inversion du signe de cette variation ; si l'entropie du sous-ensemble sélectionné augmente, la spécificité diminue. Je crois que cette façon de présenter les choses est plus claire et plus précise que celle qui consiste à identifier simplement information et nég-entropie.

* * *

Nous en avons ainsi provisoirement terminé avec la théorie de l'information et nous allons passer à la thermodynamique pour voir si, là aussi, les notions de variabilité et de spécificité sont utilisables et si elles apportent quelque clarté.

La notion d'entropie en thermodynamique n'a jamais été très facile d'accès. Poincaré disait d'elle que c'est une notion prodigieusement abstraite.

On a essayé, en vain, d'en donner des illustrations intuitives, des modèles mécaniques. Ostwald a imaginé une ingénieuse construction où toutes les formes d'énergie sont le produit de deux grandeurs : une extensité — par exemple la masse, ou le volume, ou la charge — et une intensité — par exemple le carré de la vitesse, la pression ou la différence de potentiel. L'énergie calorifique se ramènerait de même à un produit de l'extensité entropie par l'intensité température. Mais tout ceci n'est pas très éclairant, et c'est peu satisfaisant, car l'entropie n'est pas une grandeur conservative, contrairement à d'autres extensités comme la masse, ou la charge. (On m'objectera peut-être que le volume n'est pas non plus une grandeur conservative au sens strict, puisqu'on peut faire varier le volume d'un gaz alors que sa pression diminue et que son énergie reste constante. Il y a cependant une différence fondamentale entre le volume et l'entropie. Si l'on maintient la pression constante, on ne peut faire augmenter le volume et l'énergie qu'en introduisant quelque chose dans le système, ici, de la matière, du gaz par exemple. Par contre, l'entropie d'un système peut augmenter alors que ce système est rigoureusement fermé, que son énergie et sa température restent constantes. L'analogie avec les autres extensités n'est légitime que si l'on se limite aux processus réversibles, où l'entropie se conserve.)

Boltzmann a alors proposé une interprétation de l'entropie dans le cadre de la théorie cinétique.

Examinons l'exemple suivant.

On considère deux récipients, l'un plein d'un gaz parfait, l'autre vide. Et on les met en communication. Selon la thermodynamique classique, l'entropie augmente.

Boltzmann divise le volume des deux récipients en petites cellules de même grandeur. Et il étudie la répartition des molécules dans ces cellules. Il montre alors que la répartition la plus probable est celle où il y a le même nombre de molécules dans toutes les cellules, et que l'énorme majorité des complexions comporte une répartition *à peu près* égale.

Boltzmann met donc l'entropie en rapport avec la probabilité : dans le processus de la diffusion d'un gaz parfait, le système passe d'un état de répartition inégale, où les cellules de l'un des récipients sont toutes pleines et celles de l'autre toutes vides, à un état de répartition à peu près égale dans toutes les cellules des deux réci-

pients. La première répartition est très improbable, la seconde est très probable. Or la première a une entropie moins élevée que la seconde. Boltzmann a donc mis en relation l'entropie et la probabilité selon la formule

$$S = k \, l_n \, W$$

(plus exactement, c'est Planck qui a donné cette forme à la relation de Boltzmann).

Il en est de même pour la température : on divise l'espace des moments en cellules et on examine la répartition des molécules dans ces cellules (les molécules groupées dans la même cellule ayant à peu près la même direction et la même vitesse, si elles ont toutes la même masse). On montre de la même manière que, pour le volume, les complexions les plus probables correspondent à la répartition égale des molécules dans les cellules et qu'une égalisation de température entre un corps chaud et un corps froid fait passer le système des deux corps d'un état très improbable vers un état très probable en même temps que l'entropie s'accroît.

Il est évident que cette définition probabiliste de l'entropie contient une part d'arbitraire. D'une part, le volume des cellules de volume ou des cellules des moments n'est pas fixé, et la probabilité en dépend. D'autre part, on peut se demander ce que veulent dire « probable » et « improbable » dans l'optique de Boltzmann. L'état de la situation initiale (où la répartition est inégale) n'est pas probable ou improbable en soi ; si les récipients sont séparés, le système se trouve dans un état de probabilité maximum compatible avec cette situation. La situation initiale n'est improbable que dans la situation finale : il est extrêmement improbable que, *les deux récipients étant en communication* et les pressions ayant eu le temps de s'égaliser, toutes les molécules se trouvent par hasard dans l'un des deux récipients.

On voit immédiatement que l'entropie de Boltzmann est une entropie *relative*, qu'il s'agit de l'entropie de l'état initial par rapport à l'état final ou de l'état final par rapport à l'état initial, en un mot de l'augmentation de l'entropie entre l'état initial et l'état final.

Ceci nous conduit à interpréter l'entropie de Boltzmann comme une *spécificité* — puisqu'elle est relative et égale à une différence entre deux entropies ; il s'agit donc de la spécificité de l'état initial par rapport à l'état final. Mais il y a encore une difficulté, c'est

que — si l'on adopte les conventions de signes de Clausius — l'entropie *augmente* de l'état initial à l'état final. Or la spécificité de l'état initial a une certaine valeur positive, alors que la spécificité de l'état final par rapport à ce même état final est évidemment nulle ; la spécificité *diminue* donc. Il faut par conséquent inverser le signe de la spécificité : *l'entropie de Boltzmann est une spécificité négative.*

Dans cette optique boltzmannienne, l'interprétation de la quantité d'information — mesurée par la spécificité — comme une entropie négative paraît donc tout à fait justifiée.

Mais l'histoire ne s'est pas arrêtée là. Clausius avait certes commencé par définir la variation d'entropie relative à l'aide de la formule

$$dS = \frac{dQ}{T}$$

Mais il a rapidement été amené à postuler des entropies absolues, uniquement fonctions de l'état du système. Il y a été amené avant tout par le fait que l'augmentation de l'entropie se présente comme une différentielle totale, ce qui signifie qu'elle ne dépend que de l'état initial et de l'état final et non du chemin parcouru (à condition que ce chemin utilise des processus réversibles). De ce fait, il suffit de fixer arbitrairement l'entropie d'un seul état pour que l'entropie devienne une fonction univoque d'un état quelconque, un potentiel ; de même qu'il suffit de fixer arbitrairement une altitude — par exemple celle du niveau de la mer, qu'on définira comme étant l'altitude zéro — pour que l'altitude devienne une fonction univoque de chaque point du terrain. Il intervenait bien, dans cette définition de l'entropie absolue, une constante additive arbitraire, non déterminable dans le cadre de la thermodynamique classique, mais cette constante arbitraire n'était pas gênante, puisqu'il suffisait de la fixer une fois pour toutes.

Puis est venu Nernst, avec son « troisième principe de la thermodynamique » ; Nernst a établi que, lorsqu'on s'approchait du zéro absolu, l'entropie tendait vers une limite inférieure qu'elle ne pouvait dépasser.

Planck a alors fait une proposition fort raisonnable : puisqu'il suffit, pour définir l'entropie absolue d'un état quelconque E, de fixer arbitrairement l'entropie $E_0$ d'un état particulier, puisque d'autre part on sait que l'entropie s'approche d'une limite inférieure au zéro absolu, nous allons tout arranger de façon très commode en fixant à zéro l'entropie d'un système se trouvant au zéro ab-

solu. Cela fera en outre une belle symétrie entre l'entropie et la température.

On s'est donc résolument engagé dans la définition d'entropies absolues, tout d'abord pour des raisons de commodité, semble-t-il.

Mais ainsi, on s'éloignait de l'interprétation de Boltzmann, qui faisait de l'entropie une spécificité négative. On se rapprochait au contraire d'une interprétation de l'entropie comme une variabilité.

C'est encore Planck qui a franchi le pas décisif. On éprouvait, nous l'avons vu, certaines difficultés à fixer la probabilité W de Boltzmann, qui restait relative d'une part au volume des cellules choisi, d'autre part à l'état final pris comme référentiel. La mécanique ondulatoire a permis de fixer une valeur définie pour le produit du volume des cellules de volume par celui des cellules de l'espace des moments. (C'est l'une des formes qu'on peut donner au principe d'incertitude de Heisenberg : si l'on réduit les dimensions des cellules de volume, on accroît la précision de la localisation des molécules, mais on augmente en même temps l'incertitude sur leur moment, c'est-à-dire qu'on ne peut plus répartir les molécules que dans de grandes cellules de l'espace des moments.) Ceci fixait univoquement le nombre de répartitions possibles dans telle situation expérimentale macroscopique, dans tel état macroscopique. Autrement dit, la probabilité de Boltzmann devenait un rapport entre deux nombres ne dépendant chacun que de l'état correspondant, de même que l'entropie relative, la variation d'entropie, était donnée par la différence des entropies absolues (le passage du rapport à la différence s'explique ici encore par l'intervention du logarithme). Tout invitait donc à faire correspondre à l'entropie absolue le logarithme du nombre de complexions P

$$S = k\, l_n\, P$$

C'est précisément ce qu'a fait Planck, substituant à la définition probabiliste de Boltzmann une définition *statistique*, ne faisant intervenir que le nombre de complexions.

Mais cette définition correspond parfaitement, à une constante près, à la définition que nous avions donnée de l'entropie-variabilité dans le cas où les signes sont équiprobables. L'interprétation de Planck est donc une interprétation en variabilité, ce qui justifie l'emploi, par Shannon, du terme d'entropie pour la variabilité en théorie de l'information.

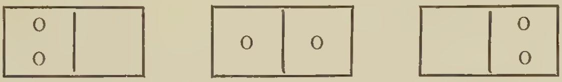
Que signifie cette nouvelle interprétation ? A quoi correspond, que représente maintenant l'entropie thermodynamique ?

L'entropie d'un état macroscopique mesure la diversité, la variété des états microscopiques compatibles avec cet état macroscopique. Par « état microscopique », nous entendons un état défini par la répartition des molécules dans les cellules du volume et de l'espace des moments. Par « état macroscopique », nous entendons un état défini par des grandeurs macroscopiques telles que masse, volume, pression, température.

Avec cette interprétation, l'entropie cesse d'être la notion prodigieusement abstraite qu'elle était en thermodynamique classique et la notion assez nébuleuse qu'elle était dans la théorie de Boltzmann. Ses caractères, ses particularités deviennent tout à fait compréhensibles.

Par exemple : lorsqu'un gaz parfait se détend dans le vide, ou lorsqu'il augmente de volume sans fournir de travail, son entropie augmente. Comment cela s'interprète-t-il ?

C'est tout simple : on lui offre de nouvelles cellules pour ses molécules. Le nombre de répartitions possibles va donc augmenter, et avec lui l'entropie-variabilité microscopique du nouvel état. Un exemple élémentaire le montrera bien. Supposons que nous ayons deux molécules et deux cellules ; le nombre de complexions possibles est trois :



Probabilité : 1/4 1/4 1/4

L'entropie, calculée en bits, est égale à 1,5.

Mais augmentons le volume en ajoutant une cellule. Le nombre de complexions passe alors à 6 :

Prob. : 1/9 2/9 2/9

1/9 2/9 1/9

L'entropie de ce nouvel état est de 2, 507 bits. Elle a donc bien augmenté.

Il en est de même pour la température : une augmentation de la température ne correspond pas seulement à une augmentation

de l'énergie cinétique moyenne des molécules, mais à une augmentation de la diversité de leurs vitesses, en valeur absolue et en direction. C'est pourquoi l'entropie du système augmente.

Enfin, le principe de Nernst s'interprète très naturellement: si l'entropie absolue est nulle au zéro absolu, cela signifie que la variété des états microscopiques est nulle, donc que le système présente toujours la même répartition de ses molécules. Et c'est évident, du moins pour la température : si toutes les molécules ont la vitesse zéro, leurs vitesses n'auront aucune variété, toutes les molécules se trouvant toujours dans la même cellule autour de l'origine, aussi petite qu'on choisisse cette cellule. Pour le volume, c'est un peu moins évident : mais du fait que le produit des cellules de l'espace des moments par les cellules de volume est fixé, si l'on peut réduire la cellule des moments autant qu'on le désire, on peut augmenter la cellule de volume autant qu'on le désire et la faire, par exemple, coïncider avec le volume occupé par le gaz. Il est clair que là encore il n'y a qu'une répartition possible, puisqu'il n'y a qu'une seule cellule.

Bref, tout ceci me paraît assez satisfaisant ; les notions de variabilité et de spécificité ont permis d'éclairer l'histoire de la notion d'entropie et son interprétation actuelle.

* * *

Nous pouvons maintenant aborder un dernier chapitre qui pourrait s'intituler : *Les principes de Carnot.*

Nous n'examinerons pas le principe de Carnot et son rôle en thermodynamique classique. Ce qui nous intéresse, c'est sa traduction en théorie cinétique, avec les interprétations de Boltzmann et de Planck.

Le principe de Carnot peut se formuler ainsi : *L'entropie d'un système fermé ne peut qu'augmenter ou, à la limite, rester constante dans des processus réversibles idéaux* ; *jamais elle ne peut diminuer.*

Comment interpréter ceci en théorie cinétique ?

Si l'on accepte l'interprétation de Boltzmann, en spécificité négative, on obtiendra un principe qui dira à peu près ceci : *Un système fermé évolue vers des états de plus en plus probables, ou de moins en moins spécifiques.*

Si, au contraire, on adopte la solution de Planck, on aura : *Un système fermé évolue vers des états de plus en plus « variables »*, « plus variable » prenant ici le sens particulier de « ayant une entropie-variabilité plus grande ».

A première vue, la première formule, correspondant à l'interprétation de Boltzmann, paraît plus évidente et par conséquent préférable : il paraît évident qu'un système évolue vers des états probables ; il l'est beaucoup moins qu'il évolue vers des états variables.

Mais il faut se méfier de certaines évidences.

Que signifie « probable » dans le sens de Boltzmann ? Nous l'avons vu : à chaque état macroscopique, à l'état initial aussi bien qu'à l'état final, correspond un certain nombre d'états microscopiques possibles. On se place dans la perspective de l'état final : les complexions correspondant à l'état initial sont improbables parmi toutes les complexions possibles dans l'état final, c'est-à-dire que le nombre de complexions a augmenté de l'état initial à l'état final. C'est précisément ce que dit la seconde formule, dans l'interprétation selon Planck. Les deux formules sont donc équivalentes, si l'on examine des choses avec soin.

Mais pourquoi la première paraît-elle plus évidente ? Elle est évidente si on lui donne le sens suivant : *Dans n'importe quelle situation expérimentale, le système se trouvera probablement dans un état probable.* Ce principe est évident, mais il ne dit pas assez. Il assure bien une évolution vers un état probable — encore faut-il pour cela négliger les fluctuations qui feraient revenir le système à un état improbable — mais il ne précise pas que l'état initial est toujours improbable par rapport à l'état final. Et c'est finalement ce qui est important.

La partie du principe qui est évidente est donc insuffisante et l'autre, qui est indispensable, se ramène à la seconde formule, de telle sorte que c'est cette seconde formule qui est finalement la plus simple et la plus directe.

Nous devons maintenant poser le problème de l'application du principe de Carnot à la théorie de l'information et éventuellement celui de l'établissement d'un « principe de Carnot généralisé », valable à la fois pour la thermodynamique et la théorie de l'information.

Je n'insisterai pas sur les difficultés qui apparaissent lorsqu'on veut faire de la quantité d'information une variabilité. En particulier, Shannon a démontré un théorème N° 7 dans lequel il établit que la variabilité d'un message ne peut que rester constante ou diminuer lors de son passage à travers un traducteur « déterministe », qu'en aucun cas elle ne peut augmenter. Malgré son analogie avec le principe de Carnot — analogie qui a séduit M. Brillouin — ce théorème dit exactement le contraire : l'évolution autorisée par le principe de Carnot correspond à une *augmentation*

de la variabilité, et non à une diminution, comme c'est le cas ici. Au contraire, la transmission avec bruit — touchée par le théorème N° 7 de Shannon — provoque en général, si la perturbation est vraiment aléatoire, une augmentation de la variabilité. On voit donc qu'il règne une extrême confusion dans les principes de Carnot de la théorie de l'information tant qu'on veut faire de la quantité d'information une variabilité.

Tout se simplifie au contraire lorsqu'on fait de celle-ci une spécificité — comme on doit le faire. Aussi bien la traduction déterministe que la perturbation aléatoire provoquent alors une diminution de la spécificité ; le principe devient alors : *Un message évolue vers des états de moins en moins spécifiques* et il est identique au principe de Carnot de la thermodynamique selon Boltzmann. Le problème d'un « principe de Carnot généralisé » semble donc avoir reçu une solution dans cette perspective.

Mais ce principe pose des problèmes. En particulier, il semble être contredit chaque fois qu'il y a sélection, choix, tri. Car alors, la spécificité augmente. Il faut expliquer comment cela est possible.

M. Brillouin l'explique ainsi : pour qu'un tri soit possible, il faut que celui qui trie acquière de l'information sur ce qu'il trie, il faut qu'il sache si la boule qu'il tient en ce moment est rouge ou blanche, pour savoir où il doit la mettre. Cette information correspond, nous l'avons vu, à une certaine spécificité. Pour acquérir cette information, il doit *voir* ce qu'il trie, et pour le voir, il doit s'éclairer. Or il ne peut s'éclairer efficacement sans utiliser des processus augmentant l'entropie, donc diminuant la spécificité. Ce qu'il gagne en spécificité macroscopique, par le tri, il ne peut le gagner qu'en acquérant de la spécificité informationnelle, cette spécificité informationnelle ne pouvant à son tour être acquise qu'au prix d'une dépense de spécificité physique. Il y aurait donc « compensation », et c'est là un phénomène bien connu en thermodynamique : l'entropie peut bien décroître dans un sous-système ouvert, mais elle croît alors dans une autre partie du système fermé total. Par exemple une machine frigorifique est capable d'abaisser la température d'un corps aux dépens de l'élévation de température d'un autre corps, ce qui représente une diminution d'entropie. Mais cette diminution se paye par une dépense de travail, dépense représentant une augmentation d'entropie qui compense et même surcompense en général la diminution obtenue. M. Brillouin postule qu'il en est de même pour la diminution d'entropie provoquée par le tri.

Je suis parfaitement d'accord sur la compensation. Mais je le suis moins sur la manière dont elle est justifiée. En effet, il peut

y avoir tri, sélection, sans qu'il y ait information : si l'on passe des cailloux à travers un crible, on sélectionne les plus fins ; il n'y a là aucune information, aucune « torche éclairante ». Et pourtant, la spécificité augmente. (On m'a objecté qu'il y avait dans le crible une nég-entropie structurale. Peut-être, mais cette nég-entropie structurale est mise une fois pour toutes dans le crible, alors que le crible peut être utilisé de façon continue. De toute façon, donc, aussi grande que soit la nég-entropie structurale, on pourra toujours rendre le bilan positif.)

D'autre part, M. Brillouin, pour résoudre le problème des copies que je ne peux aborder ici, est amené à faire une distinction très peu satisfaisante entre une information morte et une information vivante, la compensation n'intervenant, selon lui, que lors de la *lecture* de la copie, et non lors de son exécution.

Il me semble donc que la compensation doit être recherchée ailleurs, et j'ai essayé de montrer que toute mise en ordre d'un système, toute mise d'un système dans un état particulier exige une transformation d'énergie cinétique en chaleur, ou du moins d'énergie ordonnée en énergie désordonnée, ce qui représente une augmentation d'entropie. Autrement dit, *toute diminution de la variabilité quelque part dans un système, par exemple par une sélection, un tri ou autre mise en ordre, se paye par une augmentation de l'entropie-variabilité physique dans une autre partie de ce même système.* Ce principe est applicable tel quel à la compensation physique et je ne lui connais pas d'exception. Il n'est, en particulier, pas contredit par les phénomènes auxquels s'applique le théorème N° 7 de Shannon, car là aussi, la diminution de variabilité est payée par une nécessaire augmentation de la variabilité physique en quelque endroit du traducteur. C'est donc sous cette forme que je verrais un principe de Carnot généralisé, applicable également à la théorie de l'information et à la physique.

* * *

Il ne me reste plus qu'à évoquer brièvement quelques problèmes que je n'ai pu aborder, mais qui ont un rapport assez étroit avec mon sujet.

Tout d'abord, il est clair que l'application des notions de variabilité et de spécificité ne se limite pas à la thermodynamique et à la théorie de l'information. On peut, par exemple, parler de la variété, de l'entropie du style d'un auteur ; c'est une mesure de sa richesse et de sa diversité. On peut parler également de l'entropie

d'une société, d'un groupe social ; par exemple, une société libérale, individualiste, aura certainement une entropie plus grande qu'une société dirigiste et égalitaire. L'entropie peut même servir à caractériser la liberté laissée à l'individu ou du moins la liberté dont il fait usage.

La notion de *spécificité* est également utilisable dans la vie de tous les jours : certains objets, certaines structures n'ont une valeur pour nous que s'ils se trouvent dans un état particulier parmi tous les états qu'ils peuvent prendre. Ceci s'applique naturellement au message : un message n'a pour nous de la valeur que s'il se trouve dans un état bien précis, mais il en est de même pour un tableau, une mosaïque, une mélodie. Si l'on perturbe ces structures, elles perdent en même temps leur spécificité et leur valeur. Mais il ne faudrait pas en conclure que toute structure, par le fait même qu'elle est spécifique, acquiert de la valeur. L'originalité n'est pas une valeur en soi ; c'est au contraire la valeur qui, du fait qu'elle est liée à des états bien particuliers, impose une certaine spécificité. Comme il faut se garder de confondre variabilité et spécificité, il faut se garder de confondre spécificité et valeur. Il faut d'ailleurs également être prudent dans l'identification de la spécificité avec la quantité d'information ; on peut attribuer à un message une certaine spécificité sans qu'il ne transporte ou conserve aucune information.

Les structures spécifiques obéissent également à un principe de Carnot, le même que celui de la théorie de l'information, qui prescrit que la spécificité d'une structure abandonnée à elle-même ne peut que diminuer. Et ici encore se posent des problèmes au sujet de l'évolution biologique d'une part, de la création artistique, technique ou scientifique d'autre part, qui toutes deux augmentent la spécificité. Cette augmentation de spécificité est-elle de même nature que celle du choix ? Nous ne pouvons naturellement pas aborder ces problèmes ici et je renvoie ceux que cela pourrait intéresser à mon ouvrage *Information, Thermodynamique, Vie et Pensée* publié chez Gauthier-Villars.

* * *

Et, pour terminer, je vais résumer brièvement les points principaux de ma conférence.

J'ai tout d'abord exposé la théorie de l'information de façon superficielle, comme on le fait quelquefois, en faisant de l'entropie-variabilité elle-même une mesure de la quantité d'information. J'ai montré ensuite à quels paradoxes cette identification condui-

sait. On est donc amené à définir une autre grandeur, différence de deux entropies, la spécificité ; cette spécificité est la véritable mesure de la quantité d'information. J'ai montré en passant comment on pouvait interpréter cette quantité d'information soit objectivement, soit subjectivement, objectif et subjectif ayant ici un sens bien particulier.

Dans une seconde partie, j'ai montré comment ces notions peuvent être appliquées à la thermodynamique et comment l'entropie, qui avait tout d'abord été interprétée par Boltzmann comme une spécificité négative, a reçu de Planck une autre interprétation, dite « statistique », et qui fait d'elle une variabilité.

Enfin, dans une dernière partie, j'ai parlé, en rapport avec ces notions, des principes de Carnot de la thermodynamique et de la théorie de l'information, ainsi que d'un éventuel principe de Carnot généralisé.

# Causality, consciousness and creativity

by Sid Deutsch,
*Electronics Engineer,*
*Rockfeller Institute (New York)*

## INTRODUCTION

The ultimate aim of science is to discover the laws of the universe. Scientists are driven in this direction by an evolutionary-born curiosity. Man is constantly probing the infinities of space, time, matter, and energy (and, at the other extreme, their infinitesimals).

Although the laws must be restated, from time to time, as new discoveries are made, science has progressed by tentatively assuming that the laws are inviolate. A dramatic example is provided by the calculation of the mass of the earth. Newton stated that the gravitational force of attraction between any two bodies is given by $F = Km_1m_2/d^2$, where $m_1$ and $m_2$ are the masses of the bodies, $d$ is the distance between their centers of mass, and K is a constant (the gravitational constant). The value of K is determined by actually measuring the force of attraction between two bodies that are suspended in the laboratory. Knowing K, it is a simple matter to calculate that the mass of the earth is $6 \times 10^{24}$ kilograms. Knowing the mass of the earth, it is possible to determine the masses of the moon, sun, and planets. In this way, the laws of the universe serve as stepping-stones between man and his surroundings.

Science has matured to the point where, with one exception, we recognize that the universal laws also apply to living matter. We agree that man's body is an electro-chemical engine ; its molecules obey the law of gravitation ; its electric currents flow in accordance with current = voltage/resistance ; its chemical changes obey the law of conservation of energy ; its chemical combinations obey the dictates of valence.

The one exception is the principle of causality, the principle that effect uniquely follows cause. Man's vanity has led him to believe that he is an intelligent, humane, creative piece of matter, possessed

of " free will ". If causality is applied to man, however, he appears to be a complicated electro-chemical structure with no more " free will " than an atom of radium has when it " decides " to disintegrate. In a sample of radium that contains a large number of atoms, half of the atoms will disintegrate in 1600 years. We may not know why individual atoms disintegrate, but we can be certain that a definite combination of circumstances *causes* this *effect* each time, without exception. In the case of man, every electron that moves, every thought that occurs, every word that is written, must be the inevitable and predictable effect of the almost-infinite number of causes and effects that have preceded it since the beginning of time.

From an atomic point of view, the human brain is, indeed, a vast empty space differing very little from a block of wood. Here and there a nucleus appears, with an electron far off in the distance. This is a world that is devoid of free will. Although its future is predictable, the macroscopic structure is so complicated that only a few gross effects can be foretold with any reasonable probability of success.

If free will is an illusion, are consciousness and human creativity also non-existant abstractions ? In order to answer this question I will present a plausible model of the human brain. The model obeys the laws of the universe, of course. It can be constructed out of brain tissue or via the usual assortment of resistors, capacitors, inductors, transistors, batteries, and diodes. In principle, we can build a robot that will be intelligent, humane, and creative ; we can even build an artificial brain that will be endowed with " an awareness of being " or consciousness.

It is well to preface any discussion of the brain with a reminder that its workings are hidden in a domain of atomic dimensions. Molecules of water and atoms of sodium and potassium all have about the same diameter- roughly $3 \times 10^{-8}$ centimeter. The human sperm has a diameter of about $3 \times 10^{-4}$ centimeter, so that a sperm contains the equivalent of $10^{12}$ molecules of water. If we assume that a single bit of hereditary information requires a molecule whose volume is 1000 times as large as a water molecule, the sperm can carry $10^9$ or one billion bits of information. The intricacies of the human brain, including thousands of built-in behavioral patterns, must be carried within the sperm's hereditary structure.

A single fiber in the brain can store memory information in the form of a sequence of water molecules and sodium or potassium ions. If so, the storage of one billion bits of information requires a fiber that is 30 centimeters long. If 10 bits of information are stored

each second for 16 hours a day, it takes 5 years to record one billion bits.

One more point: the voltage generated by a battery is independent of the size of its electrodes. A single sodium and potassium ion pair can generate 0.21 volt, and this happens to be the order of magnitude of observed nerve impulses.

Because of its sub-microscopic structural details, very little is known about how the brain functions. My plausible model is almost completely hypothetical. To avoid the repetitive use of such words as " possibly " and " perhaps ", the model will be presented as a fait accompli although it may in fact turn out to be the result of some bad electrical connections.

## PATTERN RECOGNITION MODELS

The hypothetical brain is, in the main, an assemblage of pattern recognition units. Sight, sound, smell, taste, touch, and thought signal patterns are stored in appropriate memory fibers. Incoming signal patterns are compared with the stored patterns. Recognition pulses are produced whenever incoming and stored patterns are similar but not necessarily identical. The recognition pulses trigger various motor and other responses.

An oversimplified pattern recognition unit is depicted in Fig. 1(a). It consists of a single memory fiber and, in close proximity, a single scan fiber. In coming signals are assembled, at the left, in a temporary storage fiber. The latter is periodically excited by clock pulses. When a clock pulse arrives, the temporary storage pattern discharges into the scan fiber, where it rapidly propagates to the right and is compared with the stored memory patterns. The discharge is non-destructive ; that is, the temporary storage pattern is not erased when discharge occurs. Instead, it shifts one storage element distance to the right when the clock pulse ends, thus making room for the next incoming signal.

To illustrate with a simple example, suppose that the clock frequency is 10 pulses per second, that Morse code letters are stored in the memory fiber, and that the letter U (· · —) is received by the temporary storage fiber. A full second is required to receive the dot-dot-dash. Its relatively slow progression along the temporary storage fiber is pictured in Fig. 1(b). The patterns that are shown here are also the scan fiber signals that are released during the clock pulses. As the comments in the right-hand column

of Fig. 1(b) indicate, the letter U is first tentatively recognized as an E, then as an I, and then as an S. Finally, 0.4 second after the letter has ended, it is definitely recognized as a U.

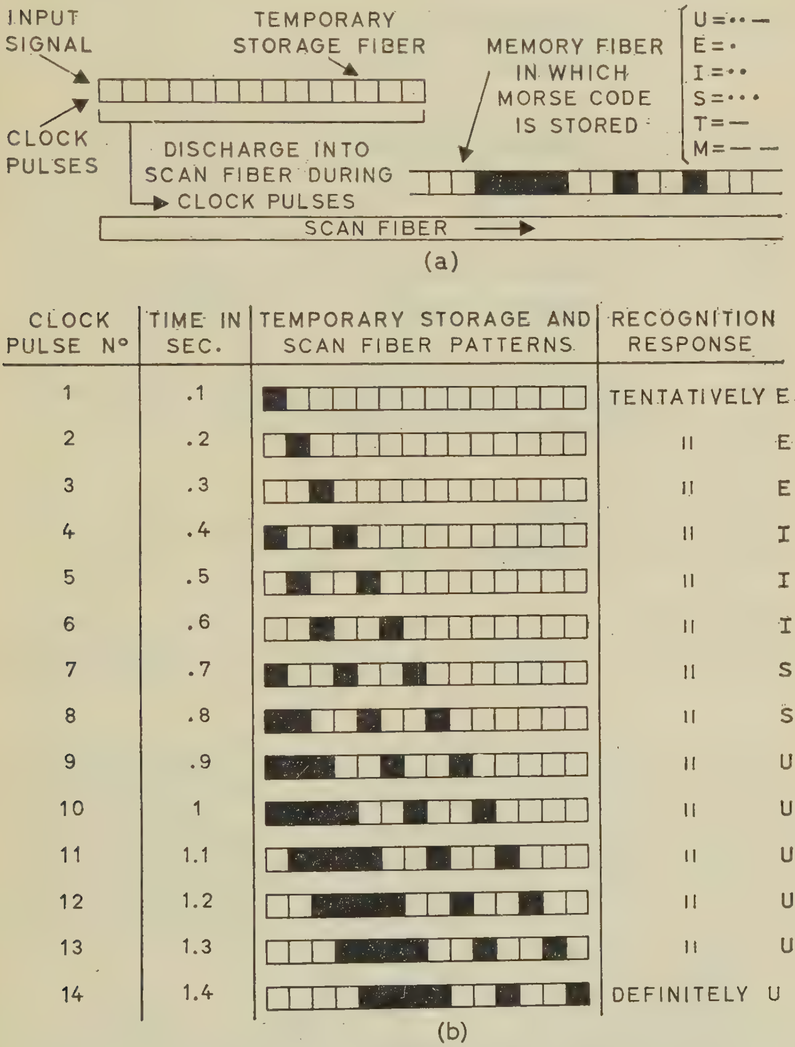



FIG. 1. — A simple pattern recognition unit for Morse code letters. *(a)* The physical layout. Incoming signals are assembled in the temporary storage fiber. They are non-destructively discharged into the scan fiber during clock pulses. The scan fiber pattern is then compared with the pattern that is stored in the memory fiber. *(b)* Temporary storage and scan fiber patterns, versus time, for a U(· · —) input signal.

Some of the requirements of a pattern recognition unit are illustrated in Fig. 2. In each case, a particular scan pattern is compared with a U memory pattern.

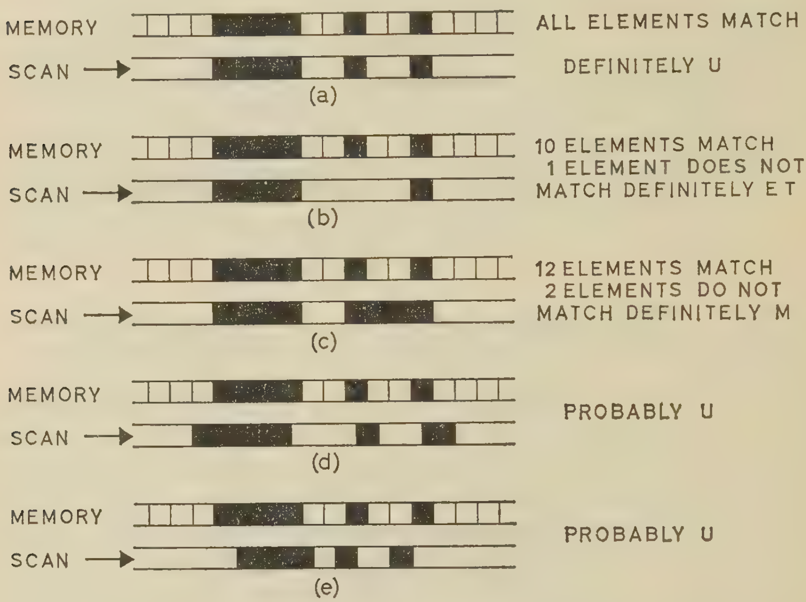



FIG. 2. — Various Morse code scan patterns compared with a U memory pattern. The desired response of the pattern recognition unit is given at the right. The following scan patterns are illustrated : *(a)* dot-dot-dash *(b)* dot-space-dash *(c)* dash-dash *(d)* a slow dot-dot-dash *(e)* a fast dot-dot-dash.

In (a), a U pattern moves along the scan fiber, and is recognized as such as it passes the memory pattern. In (b), however, the second dot in the scan U is missing. Here 5 pairs of memory and scan elements are matched while a single element — the missing dot — is unmatched. This must be sufficient to inhibit recognition as a U. The scan pattern will be recognized by other sections of the memory fiber as an E followed by a T.

In (c), the space between the dots in the scan U is filled in to form a dash. Now there are 6 pairs of matched elements and 2 unmatched elements. The scan pattern should be rejected as a U but it should receive recognition, instead, wherever M is stored in the memory fiber.

A slow 1.2-second scan U is displayed in (d) while a fast 0,8-second U is shown in (e). The scan pattern must be recognized as a probable U in either event.

From the above, it would appear that recognition must be inhibited when minor differences in *shape* exist between the scan and memory patterns, but recognition should occur when minor differences in *size* exist. These requirements can be met if the memory pattern is a source of, let us say, positive voltage while the scan pattern is a source of negative voltage. If the fibers are embedded in a conducting medium, current will flow between the memory and scan voltage distributions.

Figure 3(a) depicts a U memory pattern versus an M scan pattern. As the current flow lines demonstrate, practically no current flows to the mid region of the right-hand M dash ; this can serve as a basis for non-recognition. By contrast, Fig. 3(b) shows two patterns that have the same shape but different sizes ; here the current flow lines are fairly uniform.

A system that can distinguish between the two extremes is portrayed in Fig. 3 (c). The memory and scan fibers are surrounded by ring-shaped sensing elements.

Each of these is governed by the following three rules :

— If the enclosed region of the fiber is neutral, the sensing element is inactive.

— If the enclosed region is a source of voltage, and also supplies a relatively large amount of current, the sensing element generates a " recognition " pulse.

— If the enclosed region is a source of voltage, but supplies very little current, the sensing element generates an " inhibit recognition " pulse.

To inhibit recognition when only minor differences in shape exist, one must suppose that a single inhibit pulse is equal and opposite to approximately 10 recognition pulses. The sensing elements form a tree-like array, as shown in Fig. 3 (c), to facilitate local comparisons between large and small amounts of current. In addition, this permits the recognition of some regions of a pattern even though other regions are dissimilar.

The above discussion has concentrated on the longitudinal aspects of sensory patterns. Equally important is their transverse distribution. Consider, for example, a cross-section through a visual channel that contains 81 memory- scan fiber pairs, as in Fig. 4. The black squares symbolize a symmetrical letter A that is stored in the memory fibers. The white squares represent a distorted A that is propagating through the scan fibers. Sensing element rings have been omitted for the sake of clarity. If we regard the black and white

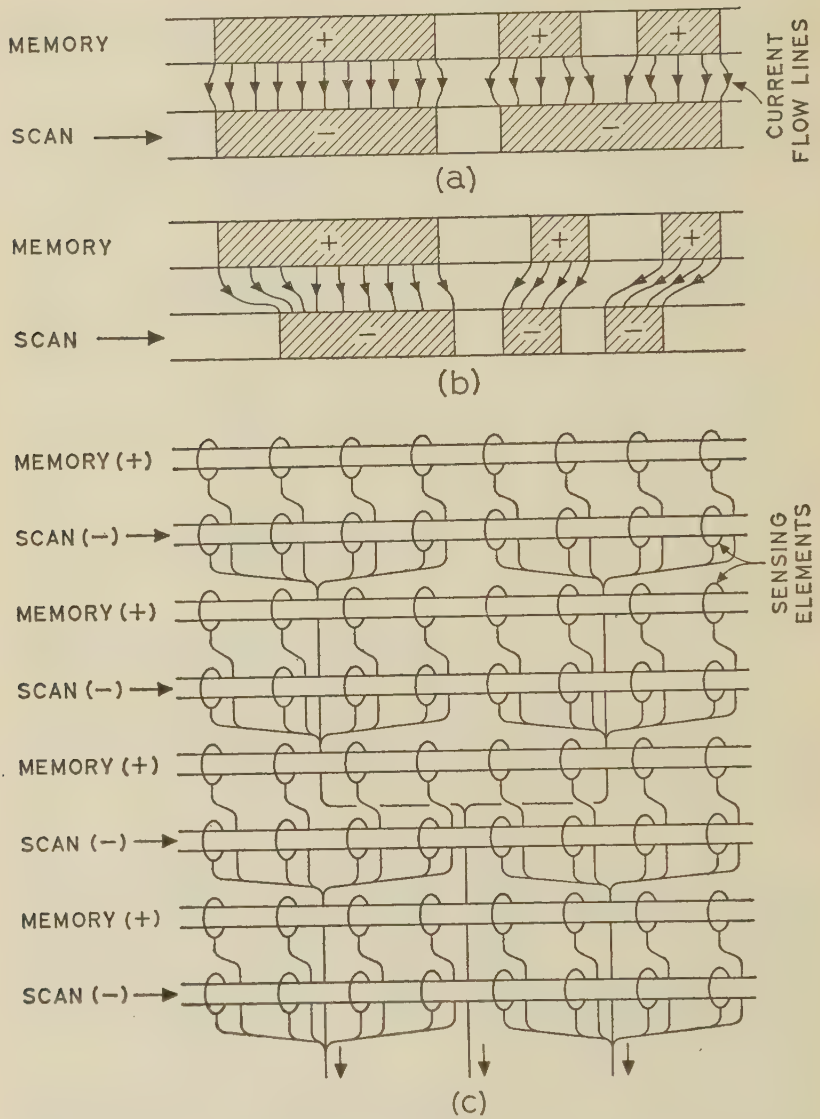



FIG. 3. — The electrical details of a pattern recognition unit. The memory pattern is a source of positive voltage while the scan pattern is a source of negative voltage. The fibers are embedded in a conducting medium. *(a)* A U memory pattern versus an M scan pattern. *(b)* A normal U memory pattern versus a fast U scan pattern. *(c)* The assembly of memory fibers, scan fibers, and sensing elements.

squares as sources of positive and negative voltage, respectively, then appreciable current will flow between all active fibers except along the upper left edges of the As and the single memory fiber in the lower left corner. The sensing element system should be adjusted so that the inhibit pulses are not quite sufficient to overcome the recognition pulses in this case.

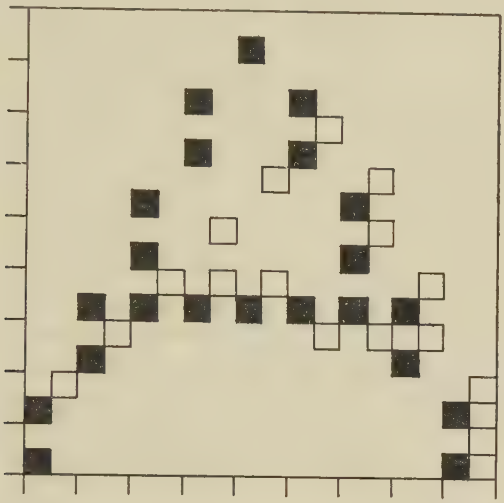


FIG. 4. — A cross-section through a visual pattern recognition channel that contains 81 memory-scan fiber pairs. The black squares symbolize a symmetrical letter A that is stored in the memory fibers. The white squares represent a distorted A that is propagating through the scan fibers. Sensing element rings have been omitted for the sake of clarity.

It would be naive to suppose that the fibers in an actual brain are mere extensions of their sensory organs, as Fig. 4 implies. Evolutionary development must be such that the fibers are rearranged so as to enhange pattern recognition. This idea is illustrated, in Fig. 5, for a visual channel. The circular pattern in (a) is that of the letter K as it is seen by the eye or retina. The retinal fibers lead to memory and scan recognition channel fibers that are mapped in modified polar coordinates. That is: angular distances in the circular field are mapped into horizontal distances in the rectangular field while radial distances in the circular field are mapped into vertical distances, using a logarithmic scale, in the rectangular field. This completely distorts the original image, but it accomplishes three things. First, the central region of the retina is spread out relative to the outer regions, so that the recognition of small objects and details is

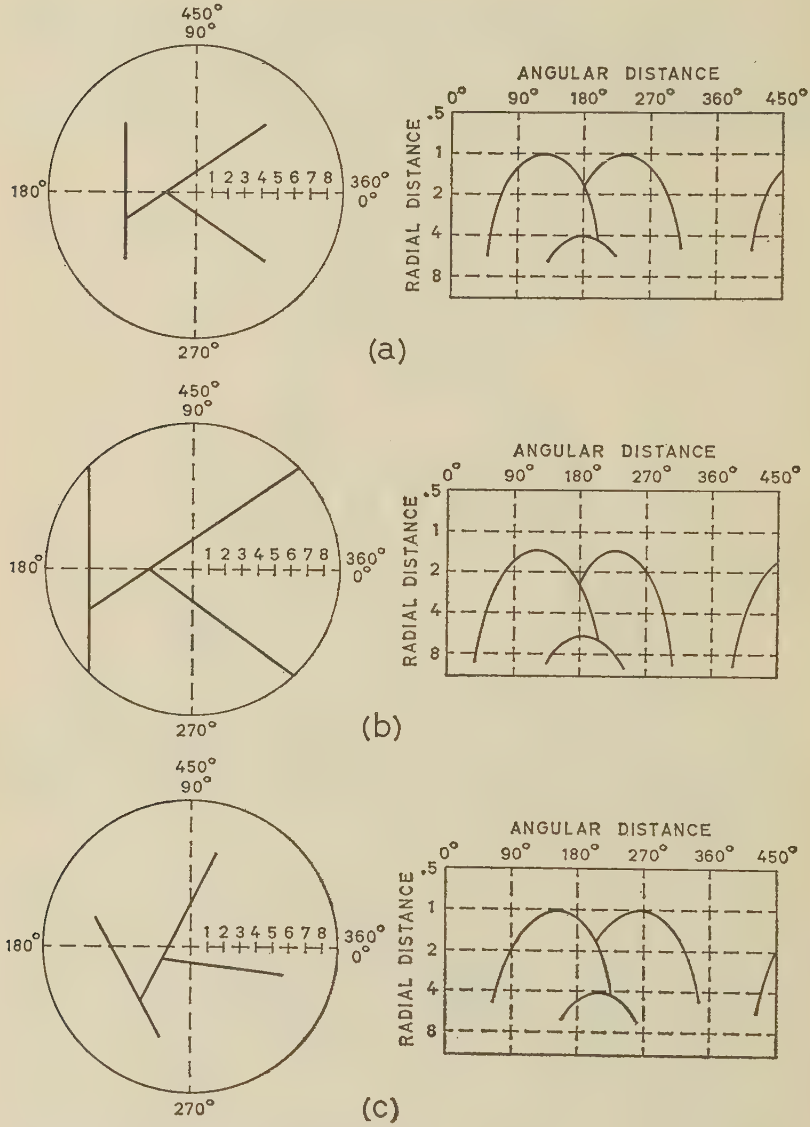



FIG. 5. — The circular pattern on the left is the retinal image of the letter K. In the rectangular field on the right, the circular field is mapped in modified polar coordinates. The rectangular field pattern size and orientation remain constant as the original K of *(a)* is magnified by a factor of 1.5 in *(b)* or rotated counterclockwise 30° in *(c)*.

enhanced. Second, a change in retinal pattern size, as in (b), results in a vertical shift of the rectangular pattern without a corresponding change in its size. Third, rotation of the retinal pattern, as in (c), results in a horizontal shift of the rectangular pattern, again without a change in size or orientation.

Similarly, in an audio recognition channel, it would be logical to arrange the fibers so that a single octave is spanned in the horizontal direction while its harmonics and sub-harmonics are measured off in the vertical direction.

A channel that duplicates the human brain must accomplish much more than " mere " recognition. When recognition occurs, the stored pattern must be circulated elsewhere to possibly awaken other associated memories. Recognition, if sufficiently intense, must also initiate appropriate motor responses. All of these features are provided in the sophisticated cross-section of Fig. 6.

In the upper portion of Fig. 6, the M circles are memory fibers while the S circles are incoming sensory information scan fibers. Patterns that have been fed back from other portions of the channel scan through the RS, or " recognition scan ", fibers. Recognition can occur between M and S fibers or between M and RS fibers.

The outer circles that surround the M, S, and RS fibers are the ring-shaped sensing elements. When recognition takes place, pulses are sent to the RO or " recognition output " fibers, as the arrows indicate. This causes the non-destructive transfer of the stored memory voltage pattern into the RO fibers.

In the lower portion of Fig. 6, the MM circles are " muscle memory " fibers. At some time in the past, when a given sensory pattern was stored in the M fibers, the corresponding pattern of muscular activity was stored in the MM fibers, also in the form of a positive voltage source distribution. The MO or " muscle output " circles are the nerve fibers that lead to individual muscles. Recognition pulses travel to the MO fibers, thereby initiating the non-destructive transfer of the stored MM pattern into the corresponding MO nerve fibers. In this way, each stored sensory memory pattern is linked to a given motor response.

Recognition pulses also enter the " time-coincident inter-sensory trigger network " bus. This refers to the other sensory and muscle patterns that were stored at the same time that the M pattern of Fig. 6 was stored. All patterns that were stored in time-coincidence are stimulated when any one of them is involved in pattern recognition. This serves to establish a catalogue of sensory stimuli and motor responses in which sight, sound, smell, taste, and touch are linked together.

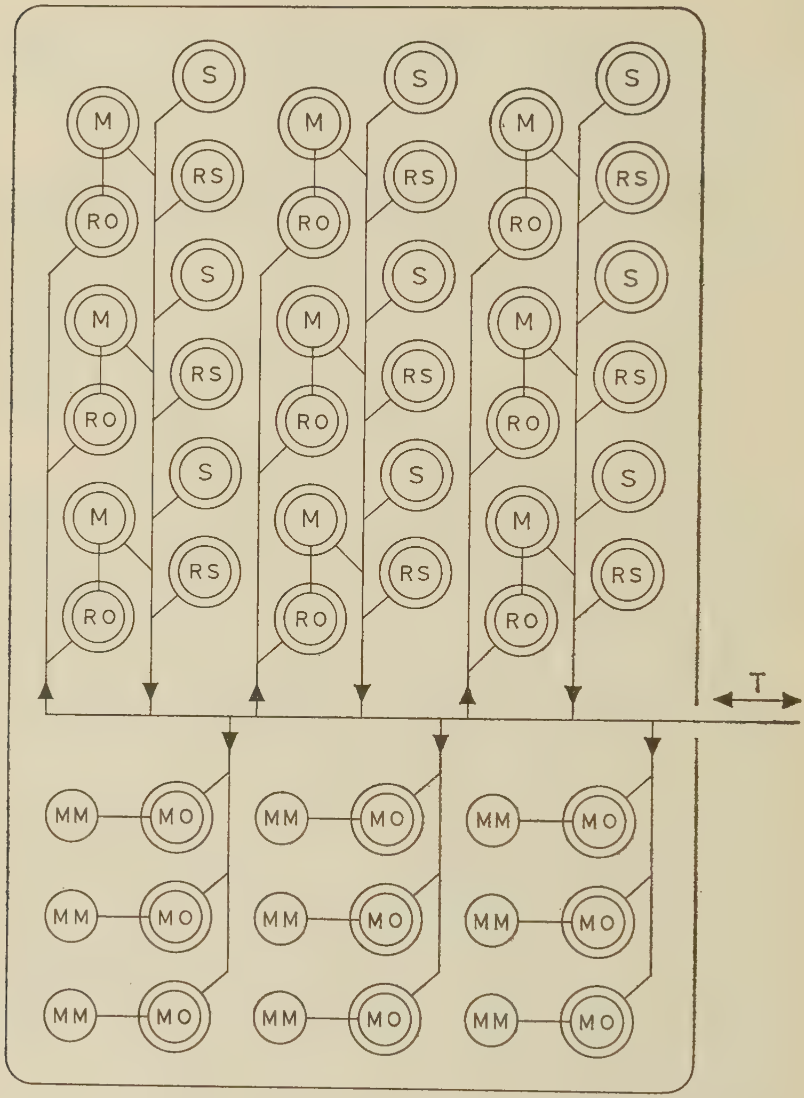



FIG. 6. — A cross-section through a sophisticated pattern recognition channel. T stands for "to and from time-coincident inter-sensory trigger network". The circles represent various fibers as follows:

M sensory memory
S sensory scan
RS recognition scan
RO recognition output
MM muscle memory
MO muscle output

## MODEL OF THE HUMAN BRAIN

Using the recognition channel of Fig. 6 as a building block, we can proceed to construct a model of the entire brain. This has been done in Fig. 7. Only the visual portion of the brain is shown, but connections to the other sensory portions are indicated.

Aside from the "Main" and "Visual Feedback" control centers, the model consists of 3 sections, in keeping with 3 types of memory storage; the instinctive, sensory, and thought memory sections.

The uppermost section is that of instinctive memory. This is akin to the brain of an insect. There is no consciousness, no awareness of pain or pleasure, no learning, no thought process. An insect's brain is minute; similarly, man's instinctive needs could be satisfied by a short section incorporated into the end of each sensory organ nerve. To illustrate: if 1000 instinctive patterns are built into a sensory nerve ending and each pattern is represented by a sequential array of 1000 water-sodium-potassium units, then a channel length of 0.03 centimeter is sufficient.

In the instinctive memory section of Fig. 7, the memory itself is symbolized by the uppermost line. Input signals from the eye scan, to the right, along the second line. The "recognition scan" line symbolizes thought patterns that are fed back from other parts of the brain. "Recognition scan" is the agent that causes the mouth to water when we think of food; these fibers are absent in the brain of an insect.

When recognition occurs, pulses are sent to the "muscle output" line via the dotted vertical connection that is associated with the recognition area. The "muscle memory" line then discharges into the "muscle output" line, thereby evoking the appropriate instinctive motor response to an external stimulus. Body regulatory mechanisms are also triggered in this manner.

The brain of a lower mammal differs from that of an insect because of the addition of the next section- that of sensory memory. Here there is consciousness and learning, but no abstract thought. This section is relatively large; as mentioned previously, one billion bits of information occupy at least 30 centimeters of fiber length. Since a single visual fiber receives much more than one billion bits of information in a man's lifetime, one must assume that the memory becomes compressed as time goes on. Where the initial memory consists of a sequential array of 1000 water-sodium-potassium units, the array may be reduced to 100 units after 5

years and 10 units after 10 years ; i.e., the fine details of memory are gradually lost. To maintain proper registration, the scan pattern must be similarly compressed as it propagates to the right.

Incoming visual patterns enter the sensory memory line and are stored in its fibers in the form of the above-mentioned sequence. To make room for new bits of information, the visual memory array slowly migrates to the right, at the same time being subjected to a gradual loss of fine detail. All of the memory patterns in the brain model migrate in synchronism ; their rate of progression is determined by the same clock pulses that govern the temporary storage periods.

The incoming visual patterns move rapidly along the " input scan " line. If the letter A is seen, for example, recognition may be triggered at the hundreds of points along the channel where similar letter As have been stored as memory. Wherever recognition occurs, the visual memory pattern is transferred to the " recognition output " line. If hundreds of As are recognized, a composite A pattern travels along the " recognition output " line and is then fed back to the " recognition scan " line. Recognition scan signals are compared with visual memory just as input scan signals are compared.

From an electrical viewpoint, the model is characterized by positive feedback. If A is recognized and fed back, it will recognize itself again and again, monopolizing the visual arena of the brain. This is prevented by depletion of the A memory sources as follows.

The electrical analogue of a pattern that is circulating around the feedback loop is generated by sodium-potassium ion batteries. These temporarily exhaust themselves in a very short time, so that new patterns tend to become dominant in the feedback loop. The new patterns are those that were associated with the previous patterns. Wherever A is recognized, for example, the visual memory patterns that are adjacent to A (i.e., patterns that occured slightly before and after A in time) are also transferred to the " recognition output " fibers. When A declines, the new patterns — hundreds of them — are ready to take over.

We must suppose the existence of a " Visual Feedback Control Center ", as indicated in the feedback loop of Fig. 7, wherein only the strongest of many signals is allowed to traverse the feedback loop. The control center is a temporary storage unit whose pattern is periodically released during clock pulses. The elimination of old signals by fatigue and the enhancement of the strongest of many new signals serve to establish a constantly-changing " stream of thought ".

Associated regions of the "muscle memory" and "muscle output" lines are triggered by each recognition pulse. The muscle memory is acquired after birth and migrates in unison with visual memory in contrast to the built-in and stationary muscle memory of the instinctive section. As shown in Fig. 7, "muscle memory" stores the patterns that appear on the "muscle output" line. The muscle memory of an infant consists solely of instinctive responses to stimuli. Later, random muscle movements occur, perhaps triggered by electrical noise in the absence of recognition triggers. Still later, the child at play builds up a muscle memory reservoir that is associated with visual memory. Finally, as a student, recognition of the letter A triggers the motor responses that cause him to write the letter A.

When visual recognition takes place, pulses are also sent to the memory fibers of all the other sensory channels, as indicated by the "to and from time-coincident inter-sensory trigger network" note. The pulses cause all memories that were stored at the same time (i.e., all events that occured simultaneously) to be transferred to each sensory "recognition output" line. Thus, if the sound A is heard very often when the letter A is seen, many coherent A sound patterns enter the sound recognition output line whenever A is seen, so that sight and sound become associated.

Feedback loops within feedback loops are set up by the branches that interconnect the various sensory memory sections. Such a system would soon have thousands of diverse thought and muscle patterns in simultaneous circulation. To avoid this epileptic catastrophe, the ten sensory and thought pattern input lines are brought together (electrically, if not physically) in the Main Control Center at the left side of Fig. 7. Here the ten patterns are compared, and only the strongest one is normally allowed to proceed without attenuation. It becomes a primary signal while the other nine patterns become secondary signals. The latter are subdued to the point where their intersensory recognition triggers are too weak to initiate feedback oscillations.

When a new and strange sensory pattern enters the model, it travels along the input scan line without inducing recognition. This alarming fact is relayed back to the main control center, which promptly regards the new pattern as the primary input signal. Secondary signals are almost completely suppressed to maximize the probability of recognizing the strange signal.

The main control center is a composite of the temporary storage fibers of Fig. 1. The primary and secondary signals discharge into

the scan lines during each clock pulse and then migrate one unit distance to the right.

Man's brain differs from that of a lower mammal in that the last section — the thought memory section — enables abstract thought

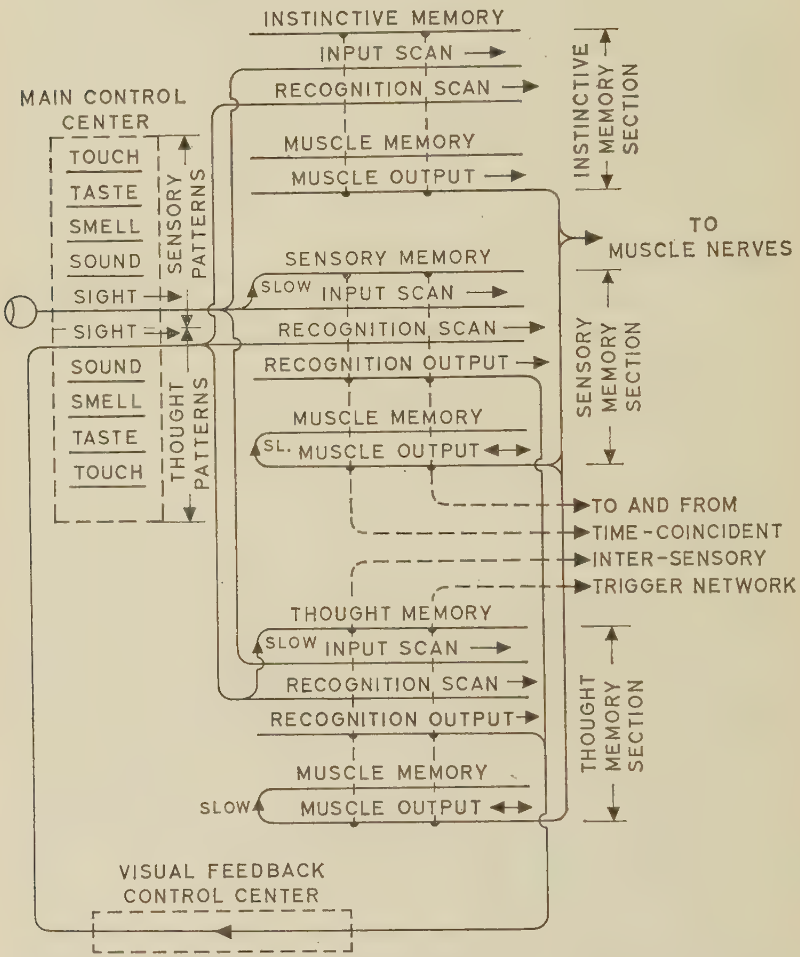



FIG. 7. — A model of the human brain. Only the visual portion is shown, but connections to the other sensory portions are indicated.

to occur. The section is identical in plan to that of sensory memory, except that thought patterns (recognition scan patterns) are stored in its memory fibers. The model implies that abstract thought is the ability to recognize a thought pattern that occured, for example, a

week ago. The thoughts of an " unintelligent " animal are always related to its immediate sensory stimuli. The animal's mind is almost completely blank when it is asleep ; it does not have dreams or nightmares. A man, on the other hand, " day-dreams " as soon as the immediate sensory stimuli become unimportant. The thought memory section is also active during night-time dreams and during hallucinations.

According to the model, intelligence involves the ability to remember previous thoughts accurately and permanently, the ability to recognize incoming stimuli or thought patterns against the reservoir of previous thoughts, and the ability to carry on a rapid chain-reaction in which one thought follows another related thought.

## CONSCIOUSNESS AND CREATIVITY

If Fig. 7 is a plausible model of the human brain, what are consciousness and creativity in the light of causality ?

Consciouness is an awareness of the ten patterns that are temporarily stored in the Main Control Center. A characteristic of consciousness is the inability to concentrate on more than one thing at a time. The main control center decides which of the ten patterns is " strongest " or " strangest " and attenuates the other nine patterns to prevent feedback instability. The ten patterns, in whatever form temporary storage is actually implemented, are the physical embodiment of consciousness.

Human creativity is the random generation of new thought (recognition output) patterns. These are random in the sense that large-population statistics are valid, but not in the sense that cause and effect are unrelated. Man can create in the same way that a lightning stroke creates heat, or evolution creates new species. In fact, human creativity is an exact parallel of evolutionary creativity. In evolution, there is the random generation of new characteristics. Most mutations do not succeed because they are incompatible with maximum reproducibility. Some mutations are more successful than prior forms, and new species are born. Similarly, most thought mutations decay in the feedback loops because they lead to a weak recognition sequence. Occasionally, a new thought leads to a hitherto unknown but strong sensory and /or thought memory recognition sequence. When the new ideas are communicated to the sensory organs of other people, they trigger an explosive sequence of new patterns and the originator is hailed as a genius.

Thought mutations can be caused by electrical noise voltages that arise from thermal agitation. The recipe for creativity is intelligence (in at least one area) and also the ability to generate new thought patterns because of a susceptibility to disturbing influences such as radiation or thermal noise. In short, creativity, like free will, is a vain illusion insofar as it violates causality.

## CONCLUSION

It is fascinating to observe the forms and motions that electro-chemical engines undergo in man and to realize that these are the inevitable and predictable effects of an almost-infinite number of preceding causal relationships. Life is evolving on an almost-infinite number of distant worlds, ruled by the same universal laws that are familier to us. These distant worlds could reveal man's past and also his future. It is quite possible that the evolutionary end of all intelligent electro-chemical engines is self destruction, and that this is why non-random signals from outer space are lacking. While it is true that every facet of man's future is written in the present state of the universe, the record is too complicated for us to read. Despite the absence of reassuring messages from outer space, however, each of us must do what he can to prevent the premature destruction of man by the microminds in his midst.

Cybernetica (Namur)
Vol. IV — N° 3 — 1961

# Théorie microscopique de l'information

par E. Scano,
*Ingénieur E.S.E.*

### Information et connaissance

L'information traitée sous la forme logarithmique habituelle

$$I = K \operatorname{Lg} \frac{P}{P_0} \tag{1}$$

est un élément quantitatif, lié au rapport des possibilités P et $P_0$ : deux informations $I_1$ et $I_2$ correspondant à deux rapports de possibilités

$$I_1 = K \operatorname{Lg} \frac{P}{P_0}$$

$$I_2 = K \operatorname{Lg} \frac{P'}{P'_0}$$

sont égales lorsque les rapports sont égaux, soit

$$I_1 = I_2$$

pour

$$\frac{P}{P_0} = \frac{P'}{P'_0}$$

De cette définition, il ne se dégage aucune notion de valeur. Le fait de dire que les informations sont égales lorsque les rapports de possibilités sont égaux, exclut implicitement la valeur attachée au résultat obtenu par l'information ; ce résultat est en somme la limitation de la liberté à P possibilités sur les $P_0$ initiales pour $I_1$ et P′ possibilités sur les $P'_0$ initiales pour $I_2$.

Le télégraphiste, dont la mission est de transmettre de l'information, est surtout intéressé par la capacité des voies de transmission et la fidélité du résultat ; la valeur de l'information

transmise lui importe peu. Le destinataire de l'information fait son affaire de la valeur qu'il doit y attacher. Il est certain que cette valeur n'est pas identique pour chaque destinataire, mais elle existe.

Le professeur qui enseigne est lui aussi un télégraphiste : il utilise bien entendu des voies différentes, mais contrairement à ce dernier, il doit s'assurer de la valeur du résultat obtenu. Si l'auditoire a une formation primaire, l'information pourra concerner un cours d'algèbre élémentaire ; par contre, si l'auditoire est dans une salle de Faculté, il pourra être question de l'axiome du choix de la théorie des ensembles ; dans l'un et l'autre cas, les quantités d'information émises, bien que pouvant être les mêmes, n'ont certainement pas atteint le même résultat. Les connaissances correspondantes n'ont rien de comparable. L'information en Faculté aura permis d'atteindre un degré beaucoup plus élevé de connaissance des mathématiques.

L'individu destinataire de l'information prodiguée par le professeur voit sa connaissance augmentée, mais il lui est indispensable d'avoir d'abord reçu l'information concernant l'algèbre élémentaire avant de recevoir celle qui concerne la théorie des ensembles ; faute de quoi, la valeur de la connaissance escomptée en faculté est loin d'être atteinte.

La connaissance apparaît donc comme la valeur attachée à l'information, les deux notions de connaissance et d'information étant liées à un individu bien défini.

L'auditeur de l'exemple précédent qui suit les cycles d'information successifs jusqu'en Faculté voit sa connaissance augmenter continuellement et au fur et à mesure de la réception d'information. Plus les quantités d'information s'accumulent au cours du temps, plus la connaissance de l'intéressé augmente. La connaissance atteinte à un certain stade de l'avancement paraît proportionnelle à la quantité totale d'information reçue depuis le départ à connaissance nulle.

Le professeur qui diffuse de l'information à destination de tiers, a déjà subi les cycles successifs et atteint une certaine connaissance à laquelle correspond un volume global d'informations ; son rôle est d'essayer de faire atteindre cette connaissance à ses élèves en diffusant de l'information. Toutefois, la quantité d'information diffusée au cours de la carrière du professeur ne semble pas avoir de relation avec sa connaissance. Il apparaît tout au plus que le professeur possédant le plus de connaissance, est susceptible de fournir le plus de connaissance à son élève. Encore paraît-il préférable que l'écart

entre la connaissance du professeur et celle de l'élève soit le plus grand possible pour permettre une augmentation plus facile de la connaissance de l'élève.

L'auditeur considéré comme mécanisme informationnel travaille à connaissance variable pendant toute la période de formation. Tandis que le professeur, dans le cadre de ses obligations fonctionnelles, travaille à connaissance constante (ou sensiblement constante) et fournit l'information nécessaire à la croissance de la connaissance de l'auditeur. Cette information peut être répétée indéfiniment et profiter à une multitude d'auditeurs.

## Connaissance et état microscopique : connaissance ponctuelle

Qu'il s'agisse du professeur, de l'auditeur ou du télégraphiste, il y a toujours connaissance. Pour le professeur, c'est la connaissance de la science enseignée, pour l'auditeur, c'est l'acquisition de la connaissance correspondante ; quant au télégraphiste, c'est la connaissance de son art. Chacun de ces individus traite de l'information et est assimilable, dans un sens beaucoup plus large, à un mécanisme informationnel.

Ces mécanismes informationnels traitent l'information, soit à connaissance constante (professeur et télégraphiste), soit à connaissance variable (auditeur).

Lorsque l'on pense la structure d'un tel mécanisme informationnel de façon à éclairer la nature de la connaissance, il paraît souhaitable d'y associer un état microscopique. La connaissance apparaît en effet comme la résultante d'un ensemble d'informations. L'information se présente à l'analyse comme une somme d'informations élémentaires à caractère granulaire. Ainsi, plus la connaissance et par conséquence l'information sont faibles, plus on tend vers l'état granulaire.

Nous dirons donc que la connaissance est ponctuelle ; c'est-à-dire que nous appellerons connaissance le fait qu'il existe une relation biunivoque et réciproque entre l'état physique d'un certain nombre de points matériels percevant chacun une information élémentaire et celui d'un ou plusieurs points matériels du mécanisme informationnel. Si la connaissance est telle que le nombre de points matériels du mécanisme est réduit à l'unité, nous dirons que le choix issu de la connaissance a une probabilité maximum pour l'ensemble des points choisis. Par contre, si ce nombre est supérieur à l'unité, la

probabilité ne sera pas maximum pour le choix correspondant. Plus loin, nous verrons comment cette précision n'a rien à voir avec l'erreur pouvant exister par rapport à la connaissance maximum.

## Système informationnel

Un système informationnel est formé par une partie de l'ensemble des points matériels appartenant au mécanisme informationnel et définissant la connaissance. Le nombre de points matériels du système informationnel forme un sous-ensemble de l'ensemble des points matériels. Soit $n_A$ l'ensemble des points matériels appartenant au mécanisme informationnel et $n$ le sous-ensemble ; on a

$$n \in n_A$$

et $n$ sera dit le volume du système informationnel.

En l'absence d'information et à connaissance nulle, le choix issu de la connaissance ne porte sur aucun point matériel particulier ; ce choix peut donc être assimilé à n'importe lequel des $n$ points du système informationnel.

En présence d'une certaine information, le choix issu de la connaissance correspondante se porte sur un certain nombre défini de points matériels. Ce nombre est supérieur à l'unité si la précision de la connaissance n'a pas atteint sa valeur maximum. Dans le cas d'une information maximum, le choix ne se porte que sur une des $n$ positions de l'ensemble $n$. Ce choix unique et défini avec précision sera dit « choix optimum » du système informationnel correspondant aux informations perçues. Le choix optimum est lié à la connaissance de probabilité maximum. Tout autre point matériel est lié à une connaissance de probabilité moindre.

Lorsque le système informationnel reçoit de l'information, la connaissance correspondante a tendance à croître ; son fonctionnement sera dit « direct ». Si, au contraire, le système informationnel, déjà pourvu d'une connaissance, transmet de l'information au milieu extérieur, son fonctionnement sera dit « inverse ».

## Relation entre la néguentropie et l'information

L'information reçue par un système informationnel tend à réduire les possibilités de choix pour arriver à la définition du choix optimum correspondant à la connaissance maximum intéressée. Le désordre initial (à connaissance nulle), permettant un choix quel-

conque, est remplacé au fur et à mesure par un ordre croissant tendant à limiter les choix possibles. Cet ordre est mesuré par la néguentropie reçue par le système informationnel ; cette néguentropie tend à croître avec l'amélioration de la connaissance.

Soit un système informationnel de volume $n$ ($S_n$) percevant une information qui lui permet de réduire le choix possible de $n$ à $n_x$ ; l'augmentation de néguentropie du système correspond à

$$\Delta N = k \mathrm{Lg}\, n - k \,\mathrm{Lg}\, n_x$$

L'apport maximum de néguentropie est valable pour $n_x = 1$ c'est-à-dire pour une connaissance maximum correspondant au volume choisi $n$ ; on a donc

$$\Delta N_{\max} = k \,\mathrm{Lg}\, n$$

Le rapprochement des deux définitions — information et néguentropie — permet d'écrire que l'augmentation de néguentropie est proportionnelle à la quantité d'information reçue, soit

$$\Delta N = \frac{k}{K} I$$

Il semble toutefois restrictif de limiter la relation à une égalité. L'information apparaît en effet sous forme d'un flux d'informations élémentaires, perçu par le système informationnel et provenant du milieu extérieur, dans le cas de fonctionnement direct. Dans le fonctionnement inverse, l'information apparaît comme un flux d'informations élémentaires, fourni au milieu extérieur par le système informationnel possédant déjà une certaine néguentropie. L'exemple du professeur, passant son existence à diffuser de l'information à partir d'une certaine connaissance, c'est-à-dire d'une certaine néguentropie, laisse prévoir que l'information est au moins égale à la néguentropie et très probablement toujours supérieure.

L'extension de la conclusion au fonctionnement direct (réception d'information du milieu extérieur) est moins évidente, mais semble intuitivement acceptable ; l'égalité de l'information à la néguentropie suppose en effet a priori que toute information élémentaire perçue par un système informationnel apporte effectivement l'augmentation correspondante de connaissance ; il sera admis provisoirement que le fait est peu probable ; la question sera reprise plus loin.

La relation retenue est donc

$$\Delta N \leqslant \frac{k}{K} I \qquad (2)$$

## Définition de la probabilité du choix

La probabilité pour que le choix optimum se trouve parmi les $n$ possibilités du système informationnel de volume $n$ est bien entendu égale à l'unité. Il faut toutefois préciser que le choix optimum répondant à ce résultat est celui qui correspond au volume $n$. A priori, ce choix n'a pas de relation avec celui d'un autre système informationnel. Cette restriction qui paraît ici incompréhensible, prendra toute sa signification plus loin.

Si nous choisissons au hasard $n_x$ possibilités parmi les $n$ formant le système informationnel de volume $n$, la probabilité, pour que le choix optimum correspondant à $n$ soit compris parmi les $n_x$, est égale à

$$P = \frac{n_x}{n} \quad \text{avec} \quad n_x < n$$

Cette probabilité, découlant du choix au hasard qui a été fait parmi les $n$ possibilités de départ, est inférieure à l'unité et diminue au fur et à mesure que $n_x$ tend vers l'unité.

La définition du choix optimum, obtenu en réduisant au hasard les possibilités de base, ne permet donc pas d'obtenir le résultat escompté. La connaissance qui peut en résulter est atteinte avec une probabilité extrêmement faible.

Par contre, si le choix des possibilités $n_x$, au lieu d'être fait au hasard, est guidé par l'information reçue du milieu extérieur, la probabilité informationnelle ($P_i$) est différente de la précédente et a priori supérieure ; on a donc

$$P_i > P$$

Le choix $n_x$ étant le même, on a donc

$$P_i = \frac{n_x}{n_i} \qquad P = \frac{n_x}{n} \qquad \text{avec} \quad n_i < n$$

Ceci revient à dire que l'information reçue du milieu extérieur permet, en dirigeant le choix des $n_x$ possibilités, d'éliminer un certain nombre de possibilités ($n - n_i$) parmi lesquelles le « choix optimum » n'est certainement pas.

## Relation entre la néguentropie et la probabilité du choix

En reprenant les résultats précédents, nous avons

$$P_i = \frac{n_x}{n_i}$$

$$P = \frac{n_x}{n}$$

$$n_x < n_i < n$$

Posons sous forme logarithmique

$$P_i = \mathrm{Lg}\,\frac{n_x}{n_i} = \mathrm{Lg}\ n_x - \mathrm{Lg}\ n_i$$

$$P = \mathrm{Lg}\,\frac{n_x}{n} = \mathrm{Lg}\ n_x - \mathrm{Lg}\ n$$

On a

$$P_i - P = \mathrm{Lg}\ n_x - \mathrm{Lg}\ n_i - \mathrm{Lg}\ n_x + \mathrm{Lg}\ n = \mathrm{Lg}\ n - \mathrm{Lg}\ n_i$$

La différence des possibilités $\mathrm{Lg}\, n - \mathrm{Lg}\, n_i$ désigne, en fait, l'augmentation de néguentropie du système informationnel dont le nombre de possibilités est passé de $n$ à $n_i$ (avec $n_i < n$) à la suite de la réception d'une certaine quantité d'informations issues du milieu extérieur. Le choix des $n_x$ possibilités, au lieu d'être fait parmi les $n$ possibilités est fait parmi les $n_i$ possibilités définies par l'information reçue.

Par ailleurs, la différence $P_i - P$ est positive et mesure l'augmentation de probabilités du choix $n_x$ à la suite de l'augmentation de néguentropie ; si $P_i - P$ est désigné par $\Delta P$, la relation qui lie l'augmentation de probabilité à l'augmentation de néguentropie est alors

$$\Delta P = \frac{1}{k}\,\Delta N \qquad (3)$$

Bien que calculée pour un choix $n_x$, l'augmentation de la probabilité est indépendante de ce choix ; seule la néguentropie, perçue par le système informationnel, en définit la valeur.

## PRINCIPE D'ÉQUIVALENCE : ÉNERGIE — CONNAISSANCE — NÉGUENTROPIE

Soit un système informationnel de volume $n$, percevant la néguentropie maximum qui lui permet la définition du choix optimum (probabilité maximum) ; ce système est en liaison avec le milieu extérieur par un nombre P de points matériels pouvant percevoir chacun une information élémentaire.

On a

$$n = 2^{P}$$

Chacune des combinaisons de points matériels P percevant de l'information élémentaire, définit une possibilité du système informationnel. Si, pour une combinaison, le nombre de points matériels percevant une information élémentaire est égal à $\alpha$, nous dirons que l'énergie perçue par cette combinaison est égale à

$$\alpha e$$

Ceci admet que l'information élémentaire, correspondant à un point matériel tel que P, est engendrée par une énergie élémentaire $e$.

L'énergie totale nécessaire à la définition de la néguentropie maximum est alors égale à

$$E_T = e \sum_{\alpha=0}^{\alpha=P} \alpha \, C_P^{\alpha}$$

Soit $E_T = \frac{e}{2} P \, 2^P$ et $P = \frac{\text{Lg}\, n}{\text{Lg}\, 2}$. On a $2^P = n$ ; donc

$$E_T = \frac{e}{2 \, \text{Lg}\, 2} \, n \, \text{Lg}\, n$$

Posons $\frac{e}{2 k \, \text{Lg}\, 2} = s$ et $\text{Lg}\, n = \frac{\Delta N}{k}$ (néguentropie maximum). On a alors $E_T = s \, n \, \Delta N$. En posant $n \, \Delta N = C$, les relations d'équivalences s'écrivent sous la forme

$$E_T = s \, C$$
$$\frac{C}{n} = \Delta N \tag{4}$$

Nous dirons que C désigne la connaissance.

Ainsi, la valeur de la connaissance acquise C est proportionnelle à l'énergie totale, et son quotient par le volume du système informationnel choisi est égal à la néguentropie perçue.

Exprimée sous cette forme, la connaissance C, liée à l'information par la relation (voir formule 2)

$$\frac{C}{n} \leqslant \frac{k}{K} \, I$$

exprime bien la notion de valeur qui manquait à la définition quantitative de base de l'information. Pour avoir deux connaissances identiques, il faut en effet deux informations identiques, mais

aussi deux systèmes informationnels de volumes identiques. Dans ce cas, la définition du choix optimum est unique pour les deux systèmes informationnels.

Deux systèmes informationnels de volumes différents ne peuvent donc pas avoir la même connaissance : exemple du professeur et de l'élève.

## Rendement informationnel $R_1$ (apprentissage simple)

Soit un système informationnel de volume $n_1$ ayant déjà perçu une information lui permettant d'atteindre la probabilité maximum; la néguentropie correspondante est égale à $k \operatorname{Lg} n_1$ pour la définition du choix optimum.

La connaissance correspondante $C_1$ est égale à $n_1$ $\Delta N$ soit

$$C_1 = n_1 \operatorname{Lg} n_1$$

Ce système informationnel fait partie d'un mécanisme informationnel de volume global $n_r$ avec $n_r > n_1$.

En vue d'obtenir un système informationnel de connaissance $C_2$ avec $C_2 > C_1$, nous supposerons que le volume $n_1$ est augmenté d'une certaine valeur et atteint $n_2$. Le nouveau système informationnel perçoit alors l'information du milieu extérieur nécessaire à la définition de la connaissance $C_2$.

L'augmentation de connaissance $\Delta C$ est égale à $C_2 - C_1$ et représente le résultat obtenu lors du passage de $n_1$ à $n_2$ à partir d'une connaissance de départ $C_1$ ; la connaissance qu'il a fallu déployer pour obtenir ce résultat $\Delta C$ est égale à $C_2$ ; le rendement informationnel s'exprime donc sous la forme

$$R_1 = \frac{C_2 - C_1}{C_2} = \frac{E_2 - E_1}{E_2} = \frac{n_2 \operatorname{Lg} n_2 - n_1 \operatorname{Lg} n_1}{n_2 \operatorname{Lg} n_2}$$

Posons $n_2 = n_1 + \Sigma$

$$R_1 = \frac{(n_1 + \Sigma) \operatorname{Lg} n_2 - n_1 \operatorname{Lg} n_1}{n_2 \operatorname{Lg} n_2} = \frac{n_1 (\operatorname{Lg} n_2 - \operatorname{Lg} n_1) + \Sigma \operatorname{Lg} n_2}{n_2 \operatorname{Lg} n_2}$$

$$= \frac{n_1\left(\operatorname{Lg} \dfrac{n_1 + \Sigma}{n_1}\right) + \Sigma \operatorname{Lg}(n_2)}{n_2 \operatorname{Lg} n_2} = \frac{n_1 \operatorname{Lg}\left(1 + \dfrac{\Sigma}{n_1}\right) + \Sigma \operatorname{Lg} n_2}{n_2 \operatorname{Lg} n_2}$$

Pour $n_1 \gg \Sigma$ et $n_2 \gg 0$, on a

$$R_1 \simeq \frac{\Sigma\,(1 + \mathrm{Lg}\; n_2)}{n_2 \;\mathrm{Lg}\; n_2}$$

$$\simeq \frac{\Sigma}{n_2}$$

$$\simeq \frac{n_2 - n_1}{n_2} \qquad (5)$$

Pour des volumes importants, le rendement informationnel est d'autant plus faible que l'écart entre les volumes est faible.

Le résultat donné par la formule (5) exprime toutefois une valeur minimum du rendement et suppose qu'il est nécessaire au système informationnel de volume $n_2$ de recevoir à nouveau toute l'information correspondant à $C_2$ pour acquérir la même valeur. Ceci ne paraît pas indispensable. Le rendement informationnel réel est donc supérieur, soit

$$R_1 \geqslant \frac{n_2 - n_1}{n_2} \qquad (6)$$

Le rendement informationnel est égal au rendement énergétique.

### Rendement informationnel $R_2$ (auto-apprentissage)

Jusqu'à présent, il a été question d'informations reçues par un système informationnel lui permettant une augmentation de connaissance. La formule (6) montre que le rendement informationnel, lors du passage du volume $n_1$ au volume $n_2$ ($n_2 > n_1$), est positif à condition que la connaissance $C_2$ reçue par le volume $n_2$ soit effectivement supérieure à $C_1$.

Or, l'information correspondante ne peut être fournie que par un autre système informationnel de connaissance supérieure (exemple du professeur et de l'élève). Ce même système informationnel a dû lui aussi acquérir la connaissance correspondante : de proche en proche, on voit la nécessité évidente, a priori d'ailleurs, de disposer d'une source informationnelle de connaissance toujours supérieure à celle de la précédente. Cette exigence pose donc le principe de l'auto-apprentissage des systèmes informationnels.

En présence d'un problème, la connaissance acquise est maximum lorsqu'il n'existe aucune différence appréciable entre le problème intéressé et la reproduction qui peut en être faite. La connaissance est assortie d'un rapprochement entre l'information et l'effet.

Nous supposerons un groupement de trois systèmes informationnels :

— un système informationnel de volume $n(S_n)$ fonctionnant dans le sens direct et percevant l'information venue de l'extérieur ;
— un système informationnel de volume $e(S_e)$ fonctionnant dans le sens direct et percevant l'information correspondant à l'écart ;
— un système informationnel de volume $a(S_a)$ fonctionnant dans le sens inverse et fournissant à un mécanisme quelconque l'information nécessaire à la définition d'un effet.

A chacune des possibilités du système $S_n$ soit $i$, peut correspondre une possibilité du système $S_a$ soit $j$

$$i \rightarrow j$$

La possibilité $j$ de $S_a$ permet la réalisation de l'effet, dont l'information définit une possibilité $i'$ de $S_n$ et une possibilité $k$ de $S_e$.

La connaissance est maximum lorsque $i'$ est confondu avec $i$, et que la possibilité $k$ de $S_e$ occupe une position bien définie parmi les $e$ possibilités de base, cette position étant celle du choix optimum du système mesurant l'écart. Les trois possibilités $i$, $j$, $k$ (chacune dans un système informationnel) répondant à ces exigences sont telles que

— pour le système $S_n$, le choix $i$ est optimum ;
— pour le système $S_a$, le choix $j$ est optimum et donne l'effet à entreprendre : l'apprentissage de l'action est complet ;
— pour le système $S_e$, le choix de la possibilité $k$ correspond à l'écart minimum entre le résultat de l'information et celui de l'effet.

Pour toute autre possibilité du système $S_n$, on ne peut pas obtenir la correspondance de $i$ à $i'$ avec un écart minimum confondu avec le précédent $k$. S'il n'en est pas ainsi, les deux résultats correspondant à des effets différents sont identiques. Les deux effets ne peuvent être qu'identiques. Les possibilités de $S_a$ ne peuvent donc qu'être choisies identiques et ne peuvent donc correspondre qu'avec le choix de $S_n$ ; donc $i$ et $i'$ ne peuvent que correspondre avec le choix optimum de $n$, Or il n'en est rien puisque nous avons supposé que $i$ choisi, est différent du choix optimum.

Les liaisons entre les possibilités des systèmes informationnels sont donc correctes lorsque les conditions ci-dessus sont réalisées. Ces liaisons seront dites à apprentissage correct. Les connaissances correspondantes sont alors

$$C_n = n\ \Delta N_n = n \operatorname{Lg} n$$
$$C_e = e\ \Delta N_e = e \operatorname{Lg} e$$
$$C_a = a\ \Delta N_a = a \operatorname{Lg} a$$

Lorsque l'apprentissage est correct, on va montrer que $C_n = C_a$ ; en effet, l'apprentissage ne peut être correct, c'est-à-dire qu'il est impossible de trouver une autre possibilité de l'ensemble $a$, permettant la définition de $i$ et $j$ optimum, que lorsque chacune des possibilités de $n$ aura été rapprochée des possibilités de $a$ (la possibilité $j$ faisant partie de $a$). A chacune des possibilités de $n$, il est possible de faire correspondre une possibilité de $a$ telle que $i$ et $i'$ soient confondus, mais la possibilité $k$ ne peut correspondre avec le choix optimum de $e$ que dans le cas d'apprentissage correct. Donc, à chacune des possibilités de $n$, on peut faire correspondre une possibilité de $a$, soit $n = a$ et par conséquent $C_n = C_a$.

Sans apport d'information extérieure provenant d'un autre système informationnel, l'auto-apprentissage d'un système informationnel de volume $n$, pour la définition d'une connaissance $C_n$, nécessite l'apprentissage d'une action de connaissance $C_a$ avec $C_a = C_n$. Le résultat ne peut être obtenu qu'à condition qu'il existe un système informationnel d'écart de connaissance $C_e$. Cette connaissance d'écart $C_e$ peut être supposée valable pour n'importe quelle opération d'apprentissage.

Le rendement informationnel d'auto-apprentissage, correspondant au passage de $C_{n_1}$ à $C_{n_2}$ de la connaissance du système informationnel de volume $n$, lorsque $n$ varie de $n_1$ à $n_2$ ($n_2 > n_1$) est égal au quotient de l'augmentation de connaissance

$$\Delta C = C_{n_2} - C_{n_1}$$

par la connaissance totale à mettre en jeu ; cette dernière $C_T$ est égale à la somme de la connaissance nécessaire à la définition du choix optimum de $S_{n_2}$, soit $C_{n_2}$ et de celle qui correspond à la définition du choix optimum de $S_a$ ($n = a$), soit $C_a = C_{n_2}$. La connaissance $C_a$ correspond à l'apprentissage nécessaire

$$C_T = 2\ C_{n_2}$$

donc

$$R_2 = \frac{C_{n_2} - C_{n_1}}{2 C_{n_2}}$$

En rapprochant le rendement $R_2$ du rendement simple $R_1$ (formule 5), on obtient également

$$R_2 \simeq \frac{1}{2} \frac{n_2 - n_1}{n_2}$$

Pour les mêmes raisons que le rendement simple $R_1$, le rendement d'auto-apprentissage $R_2$ tel que défini ci-dessus est une valeur minimum : on a donc

$$R_2 \geqslant \frac{1}{2} \frac{n_2 - n_1}{n_2} \qquad (7)$$

### Perte informationnelle de transmission

Les résultats précédents et en particulier la relation définissant l'augmentation de probabilité en fonction de la néguentropie

$$\Delta P_{nx} = \frac{1}{k} \Delta N \qquad (3)$$

supposent implicitement que la source informationnelle extérieure possède une connaissance C pouvant être effectivement atteinte par le système informationnel récepteur de volume $n$, c'est-à-dire que les deux systèmes ont même volume ; l'un, l'émetteur, possède la connaissance

$$C = n \operatorname{Lg} n$$

tandis que le récepteur, de volume $n$ également, ne possède aucune connaissance.

Dans ces conditions, la néguentropie $\Delta N = k \operatorname{Lg} n$ correspondant au système émetteur est entièrement reçue par le système informationnel récepteur ; ce qui permet d'écrire que l'augmentation de probabilité $\Delta P$ est bien proportionnelle à cette même néguentropie fournie par le système émetteur.

Il est également supposé que les deux systèmes informationnels sont semblables, c'est-à-dire que toutes les possibilités du système récepteur peuvent être effectivement examinées par les informations issues du système émetteur ; ce qui suppose que ces mêmes possibilités ont déjà été examinées par le système informationnel émetteur possédant la connaissance

$$C = n \operatorname{Lg} n$$

Tout en continuant à considérer deux systèmes informationnels semblables, la généralisation est à examiner en prenant le cas d'un système informationnel émetteur de connaissance supérieure à celle du système informationnel récepteur, soit

— système récepteur $C_1 = n_1 \operatorname{Lg} n_1$ volume $n_1$,
— système émetteur $C_2 = n_2 \operatorname{Lg} n_2$ volume $n_2$ avec $n_2 > n_1$.

La néguentropie reçue du système émetteur $\Delta N_2 = k \operatorname{Lg} n_2$ est supérieure à la néguentropie utilisée par le système récepteur, soit

$$\Delta N_2 > \Delta N_1$$

$$\Delta N_2 = \Delta N_1 + \Delta N$$

Or, on a

$$\Delta N_2 = \frac{k}{K} I_2$$

$I_2$ est l'information totale reçue du système émetteur, tandis que $\Delta N_1$ mesure la quantité de néguentropie utilisée par le système récepteur. La relation entre la néguentropie utilisée et l'information reçue est alors

$$\Delta N_1 < \frac{k}{K} I_2$$

et vérifie l'affirmation intuitive de la formule (2).

On a également l'augmentation de probabilité du système récepteur qui est proportionnelle à la néguentropie utile

$$\Delta P = \frac{1}{k} \Delta N_u \tag{8}$$

Pour généraliser le résultat de la formule (3) au cas d'un système émetteur de volume supérieur à celui du système récepteur, mais tout en lui étant semblable, il suffit de considérer la néguentropie utile, toujours inférieure ou au plus égale à la néguentropie émise.

La transmission de néguentropie (ou d'information) d'un système à l'autre s'effectue alors avec une perte de connaissance ; la perte informationnelle est mesurée par

$$p = \frac{C_2 - C_1}{C_2} \simeq \frac{n_2 - n_1}{n_2}$$

$$p \simeq \frac{n_2 - n_1}{n_2} \tag{9}$$

pour $n_2 = n_1$. On retrouve le cas d'échange de connaissance à valeur égale ; la perte est nulle.

Le résultat de la formule (9) suppose que les deux systèmes informationnels intéressés sont semblables. Dans le cas contraire, toutes les possibilités du système informationnel-récepteur ne peuvent pas être examinées par les informations issues du système émetteur.

Posons $s \in n_1$, l'ensemble des possibilités du système récepteur effectivement examinées par les informations issues du système émetteur.

La néguentropie utilisée est alors

$$\Delta N_u = k \operatorname{Lg} s$$

La néguentropie qui aurait pu être utilisée par un système informationnel semblable est égale à

$$\Delta N = k \operatorname{Lg} n_1$$

donc

$$\Delta N_u < \Delta N$$

La perte informationnelle dans le cas de systèmes non semblables est donc supérieure à celle qui a été définie par la formule (9). L'expression la plus générale de la perte se met sous la forme suivante

$$p \geqslant \frac{n_2 - n_1}{n_2} \tag{10}$$

## Principe d'incertitude de la connaissance

Lorsqu'il a été question précédemment de définir la probabilité du choix, une restriction a été faite en disant que le choix optimum d'un ensemble de volume $n$ est une caractéristique particulière à cet ensemble. Il convient maintenant d'éclairer cette notion après avoir défini ce que l'on peut attendre des échanges d'informations entre systèmes informationnels.

Supposons les ensembles suivants :

— système $S_{n1}$ récepteur,
— système $S_{n2}$ émetteur,

$$n_1 < n_2$$

$$C_2 = n_2 \, \Delta N_2$$

Il s'agit de systèmes informationnels semblables.

Les informations reçues de $S_{n_2}$ permettent la définition d'une connaissance $C_1 = n_1 \, \Delta N_1$ du système $S_{n_1}$. Cette connaissance est caractérisée par une seule des possibilités de l'ensemble $n_1$ ; c'est le choix optimum de l'ensemble $n_1$.

Les informations élémentaires issues de $S_{n_2}$ permettent l'examen au fur et à mesure des possibilités de $S_{n_1}$, de façon à éliminer par un

choix successif les possibilités parmi lesquelles le choix optimum de $S_{n_2}$ n'existe pas.

Lorsque l'avancement de cette élimination est tel qu'il ne reste qu'une seule possibilité parmi toutes celles de $S_{n_1}$, cette possibilité sera le choix optimum de $S_{n_1}$. A priori, il n'est pas évident que ce choix optimum de $S_{n_1}$, coincide avec celui de $S_{n_2}$ dont la définition nécessite une quantité d'informations complémentaires ; posons égal à $e$ le nombre de possibilités de $S_{n_2}$ restantes.

La néguentropie fournie par $S_{n_2}$ permettant le choix optimum de $S_{n_1}$ a non seulement permis l'examen des $n_1$ possibilités de $S_{n_1}$ mais vraisemblablement, d'un certain nombre de possibilités de $S_{n_2}$, soit $d$.

On a
$$n_1 + d + e = n_2$$

La néguentropie totale est alors
$$k \operatorname{Lg} n_2 = k \{ [\operatorname{Lg} n_2 - \operatorname{Lg} (e+n_1)] + [\operatorname{Lg}(e+n_1) - \operatorname{Lg} e] + \operatorname{Lg} e \}$$

La néguentropie $k \operatorname{Lg} e$ est inutilisable par le système $S_{n_1}$ et mesure le nombre de possibilités de $S_{n_2}$ non comprises parmi les possibilités de $S_{n_1}$ et parmi lesquelles le choix optimum de $S_{n_2}$ existe ; $e$ désignera l'écart entre le choix optimum de $S_{n_2}$ et celui de $S_{n_1}$ ; on a
$$e = n_2 - n_1 - d$$
$$e \leqslant n_2 - n_1$$
$$\frac{e}{n_2} \leqslant \frac{n_2 - n_1}{n_2}$$

et en faisant intervenir la perte (formule 9),
$$\frac{e}{n_2} \leqslant \frac{C_2 - C_1}{C_2}$$

Or $C_2 = n_2 \Delta N$ (néguentropie totale de $S_{n_2}$)
$$e \leqslant \frac{C_2 - C_1}{\Delta N}$$

En posant $C_2 - C_1 = \Delta C$
$$e \, \Delta N \leqslant \Delta C \qquad (11)$$
$$e \, I \leqslant \frac{K}{k} \Delta C \qquad (12)$$

par rapport à l'information correspondant à la néguentropie de $S_{n_2}$.

Le produit de l'information par l'écart entre les deux choix optima est, au plus égal, au coefficient $\frac{K}{k}$ près, à la différence de connaissance $\Delta C$ entre les systèmes informationnels.

Ainsi, deux connaissance $C_1$ et $C_2$ sont toujours différentes si les volumes correspondants des systèmes informationnels sont différents et la plus grande des connaissances correspond au système de plus grand volume. Mais deux connaissances différentes peuvent correspondre à des choix optima, dont l'écart $e$ a une valeur quelconque satisfaisant l'une des deux relations ci-dessus (11, 12). Cet écart peut en particulier avoir une valeur nulle ; dans ce cas, nous dirons que les connaissances $C_1$ et $C_2$ sont parfaites.

## Système informationnel généralisé : étude de structure

La structure d'un système informationnel doit être telle qu'à toute combinaison des informations élémentaires perçues par les P points matériels correspond un point matériel et un seul définissant la possibilité du choix.

Une étude logique de structure [1] montre qu'il est possible d'établir des liaisons entre $n$ points matériels de façon que ces points matériels représentent chacun une possibilité de choix univoque du système informationnel, les liaisons étant entendues dans le sens habituel de possibilité de transit d'un certain phénomène physique. Quant aux points matériels, ils sont tous identiques et doivent posséder un ensemble de propriétés :

— sélectivité,
— mémoire,
— possibilité de répétition de signaux.

L'important est de noter que le nombre de points matériels ainsi définis (ou nœuds) n'excède pas le nombre de possibilités nécessaires. Il est également très intéressant d'ajouter que, pour un même nombre de points matériels tels que P (percevant les informations élémentaires extérieures), le nombre de nœuds du système informationnel peut être réduit considérablement en classant correctement les informations élémentaires perçues habituellement par un système informationnel dans l'ordre des probabilités décroissantes. Inversement, pour un même nombre de possibilités (ou nœuds), on

[1] L'étude logique de structure fait l'objet de deux documents qui seront publiés ultérieurement dans *Cybernetica*.

peut augmenter considérablement le nombre possible de points matériels tels que P et par la même le nombre d'informations élémentaires percevables simultanément.

Le résultat, non moins intéressant, est que le mécanisme informationnel est effectivement réalisé lorsque l'ensemble des nœuds identiques tels que définis plus haut, doués chacun d'un certain nombre de possibilités de liaisons, a atteint sa position la plus probable.

Le système informationnel est défini comme une partie du mécanisme informationnel. Pour permettre l'augmentation de volume du système informationnel, il suffit, dans le classement des points matériels P élémentaires, de choisir dans l'ordre décroissant des probabilités.

Pour satisfaire à la possibilité d'auto-apprentissage, il convient de supposer que les trois systèmes informationnels $S_n$, $S_a$, $S_e$, font partie d'un même mécanisme informationnel, la sélection des informations correspondantes se faisant au niveau des nœuds du mécanisme.

## Connaissance et apprentissage parfaits

Il a été vu que la connaissance parfaite est atteinte lorsque le choix optimum du système informationnel est défini par une seule possibilité de probabilité maximum, avec un écart nul par rapport au choix optimum du système émetteur.

Dans le cas général, rien ne semble rendre obligatoire l'exigence d'un écart nul. Il convient en effet de penser qu'un même système informationnel peut être amené, simultanément ou non, à pratiquer les trois opérations de base décrites plus haut, apprentissage simple, auto-apprentissage, transmission d'information. Le système informationnel émetteur (de forte connaissance) et le système récepteur (de moindre connaissance), sont à même, chacun d'eux séparément, de matérialiser la connaissance acquise en faisant usage du système informationnel $S_a$. Il est entendu que chacun des systèmes informationnels, émetteur et récepteur, est un système informationnel généralisé pouvant fonctionner simultanément en $S_n$, $S_a$, $S_e$.

Chacun de ces systèmes dispose des mêmes mécanismes capables de la définition de l'effet, soit

$S_{n1}$, système informationnel récepteur avec $S_{a1}$ et $S_{e1}$,
$S_{n2}$, système informationnel émetteur avec $S_{a2}$ et $S_{e2}$.

En reprenant les résultats acquis précédemment on peut ajouter

$$a_1 = n_1 \text{ et } a_2 = n_2$$

L'écart possible entre les choix optima de $n_1$ et $n_2$ est donc semblable à l'écart possible entre les choix optima de $a_1$ et $a_2$ et à celui qui existe entre les choix optima de $S_{e_1}$ et $S_{e_2}$.

La connaissance parfaite à écart nul n'est donc possible que lorsque le système $S_e$ du mécanisme informationnel généralisé atteint, pour les connaissances $C_1$ et $C_2$, le même choix optimum (c'est-à-dire la même connaissance).

La mesure de l'écart entre les choix optima de $S_{n_1}$ et $S_{n_2}$, c'est-à-dire entre les connaissances $C_1$ et $C_2$, est ainsi assurée par le système informationnel d'écart. Si, lors de la définition de la connaissance $C_1$ à partir de $C_2$ avec $C_2 > C_1$, la perte informationnelle de $S_e$ est égale à $p_e$ on peut écrire (formule 9)

$$p_e = \frac{e}{e_1}$$

$e_1$ étant le volume global du système $S_e$. Donc

$$e = e_1 p_e \tag{13}$$

Si la transmission de connaissance est telle que $e = 0$, la perte du système d'écart est nulle. La connaissance $C_1$ est donc parfaite lorsque l'apprentissage est parfait, c'est-à-dire que les deux systèmes d'erreurs (système émetteur et système récepteur) ont des connaissances identiques (ou des pertes nulles).

Il faut noter que $C_1$ et $C_2$ ne sont pas obligatoirement égaux. En reprenant les relations formulées lors de l'examen du principe d'incertitude de la connaissance, on voit que la différence entre les volumes des deux systèmes informationnels $S_{n_1}$ et $S_{n_2}$ est alors égale à $d$, avec $C_2 > C_1$. Il est donc possible, avec deux mécanismes informationnels de connaissances différentes, d'effectuer des actions de même précision. Le résultat contraire aurait vraisemblablement laissé planer un certain doute sur les éléments présentés lors de l'examen du même principe.

CYBERNETICA (Namur)
Vol. IV — N° 3 — 1961